Technical Report Documentation Page

1. Report No. FHWA/TX-03-1838-8	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle TRANSPORTATION CONTROL EFFECTIVENESS IN OZONE NON-ATTAINMENT AREAS: FINAL REPORT		5. Report Date August 2001 Revised July 2003
7. Authors Chandra R. Bhat, Huimin Zhao, Yasasvi Popuri, Monique Stinson, and Sara Poindexter		6. Performing Organization Code
		8. Performing Organization Report No. 1838-8
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 3208 Red River, Suite 200 Austin, TX 78705-2650		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. 0-1838
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, TX 78763-5080		13. Type of Report and Period Covered Research Report
		14. Sponsoring Agency Code
15. Supplementary Notes Project conducted in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.		

16. Abstract

The main objective of this project is to develop a prototype of an integrated and coordinated regional transportation planning and emissions modeling procedure. The procedure will be flexible in structure to accommodate continual advances in travel demand methods and improvements in the MOBILE model. The need for an integrated and coordinated TCM air quality analysis procedure has been emphasized by a review of the state-of-the-art modeling practice used by 41 MPOs and nine states in the country (Chatterjee et al. 1997). The report provides recommendations for improvements in traffic input data for emissions modeling. These recommendations are based on a detailed sensitivity analysis that shows the high sensitivity of the emissions factor model to many input traffic data.

17. Key Words Nonattainment areas, ozone, transportation control measures, emissions modeling, travel demand modeling	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.		
19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 170	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

Transportation Control Effectiveness in Ozone Non-Attainment Areas: Final Report

by

Chandra R. Bhat, Huimin Zhao, Yasasvi Popuri,

Monique Stinson, and Sara Poindexter

Research Report Number 1838-8

Project Number 0-1838

Transportation Control Effectiveness in Ozone Non-Attainment Areas

Conducted for the

TEXAS DEPARTMENT OF TRANSPORTATION

in cooperation with the

U.S. DEPARTMENT OF TRANSPORTATION

Federal Highway Administration

by the

CENTER FOR TRANSPORTATION RESEARCH

Bureau of Engineering Research

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2001
Revised July 2003

DISCLAIMERS

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data present herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration or the Texas Department of Transportation. This report does not constitute a standard, specification, or regulation.

There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine, manufacture, design or composition of matter, or any new and useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

NOT INTENDED FOR CONSTRUCTION,

BIDDING, OR PERMIT PURPOSES

Dr. Chandra R. Bhat

Research Supervisor

ACKNOWLEDGMENTS

The authors acknowledge the support of TxDOT Project Director William Knowles of RTI; George Reeves of TxDOT's Dallas District; Wayne Young of Environmental Affairs Division; Carol Nixon of TxDOT's Houston District; and Ken Cervenka, Mahmoud Ahmadi, and Chris Klaus of NCTCOG for providing documents and data.

Research performed in cooperation with the Texas Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE - INTRODUCTION

Transportation planners, traditionally used to focusing on regional travel forecasting, now have the added responsibility of providing traffic data to state air quality agencies for use in mobile source emissions analysis. Coordinated efforts among land use planners, transportation planners, and air quality planners are needed to ensure the provision of safe and efficient transportation systems while also addressing environmental concerns. In particular, the 1990 Clean Air Act Amendments (CAAAs) require states to meet national ambient air quality standards (NAAQs) for ozone, carbon monoxide (CO), and particulate matter (PM-10). Regions not attaining the standards are designated as non-attainment areas by pollutant type. Such designated non-attainment areas have specific requirements to work toward achieving attainment by a deadline date determined by the severity classification of the non-attainment.

Atmospheric emissions may be classified into three categories: mobile, stationary, and area sources. Among these, mobile source emissions contribute significantly to the ozone and CO pollutants. Approximately, 50% of ozone precursor emissions (volatile organic compounds and oxides of nitrogen – VOCs and NOx, respectively) originate from mobile-source emissions, and about 90% of CO emissions come from mobile-source emissions.

The significant contribution of mobile-source emissions to air pollution has led to the use of transportation control measures (TCMs) as an important component of an overall strategy to reduce pollution levels. TCMs strive to promote higher efficiency of the roadway infrastructure by promoting the effective use and management of the transportation system. The TCMs included in the Texas State Implementation Plan (SIP) fall into 11 categories, which may be further grouped into four broad classes: a) projects improving speed (intersection improvements, signal timing and progression projects, motorist assistance/incident detection and response /freeway surveillance), b) projects reducing vehicle miles of travel (VMT) (high occupancy vehicle lanes, park-and-ride lots, pedestrian/bicycle facilities, commuter and light rail, and travel demand management options such as carpooling incentives, telecommuting, van pooling incentives, and existing transit service improvements), c) Capacity improvements (arterial street widening), and d) technology improvements (reformulated gasoline, alternative fuels, etc.). The ones that have direct mobility impacts are the first three TCM classes.

The CAAA and the 1991 Intermodal Surface Transportation Efficiency Act (ISTEA) emphasize TCMs as a strategy for integration into transportation/air quality planning. Many metropolitan planning organizations (MPOs) and state departments of transportation rely on the TCM-induced mobile-source emissions reduction to conform to emissions budgets established in the state implementation plans (SIPs) by state environmental agencies. Such a conformity analysis requires the evaluation of TCMs in terms of mobility, emissions, and cost. An evaluation procedure for TCM impacts is also needed for funding under the ISTEA Congestion Mitigation and Air Quality (CMAQ) act.

The main objective of this project is to develop a prototype of an integrated and coordinated regional transportation planning and emissions modeling procedure. The procedure will be flexible in structure to accommodate continual advances in travel demand methods and improvements in the MOBILE model. The need for an integrated and coordinated TCM air quality analysis procedure has been emphasized by a review of the state-of-the-art modeling practice used by 41 MPOs and nine states in the country (Chatterjee et al., 1997). The report provides recommendations for improvements in traffic input data for emissions modeling. These recommendations are based on a detailed sensitivity analysis that shows the high sensitivity of the emissions factor model to many input traffic data.

The rest of this report is organized as follows. Chapter Two discusses the travel demand modeling improvements in the current study. These include: (a) an ordered response probit model for trip production, (b) a disaggregate attraction end choice model for trip distribution, and (c) departure time choice models. Chapter Three presents the traffic models developed in the project for improved forecasting of traffic variables needed as input by the MOBILE6 emissions factor model. The traffic models include VMT mix models, soak-time duration models and trip duration models. Chapter Four describes in detail the implementation and integration of the estimated travel demand and emissions models in TransCAD. Finally, Chapter Five summarizes the findings and concludes with a recommendation for implementation of the proposed work.

CHAPTER TWO - TRAVEL DEMAND MODELING IMPROVEMENT

2.1 INTRODUCTION

In this study, a four-step sequential modeling process is developed on the Dallas-Fort Worth ozone non-attainment region. The Dallas-Fort Worth area was chosen as the study area for the project because it was recently classified as a serious ozone non-attainment area. The metropolitan area for the Dallas-Fort Worth Regional Travel Model (DFWRTM) includes about a 4,980 square-mile area with about 45,000 roadway links. It includes all of Collin, Dallas, Denton, Rockwall, and Tarrant counties and portions of Ellis, Johnson, Kaufman, and Parker counties.

The model process developed in this study uses the zone structure hierarchy that is currently used by North Central Texas Council of Governments (NCTCOG). The model structures adopted in the study are consistent with the nature of the dependent variables characterizing trip making behavior. In trip generation, the ordered response choice model structure is employed to determine the total number of trips produced from each zone. The trip productions then serve as inputs to a disaggregate attraction-end choice model which determines the fraction of productions from each zone attracted to each of the other zones. The output at the end of the disaggregate attraction-end choice model, i.e., the number of production-attraction interchanges between each pair of zones, is further converged to origin-destination (O-D) matrices. The zone-to-zone O-D matrices are then split by travel mode based on the mode split model estimated by NCTCOG. Finally, a departure time choice model estimated in the project is applied to obtain O-D interchanges by mode in different time periods.

Two major inputs for the travel demand model system are the base year data and a zone structure and network system in the study area. The base year data used to develop the models include the 1996 household activity survey, the zonal socioeconomic data including the numbers of employment by type for each zone, the zonal land use data containing acreage information of various facilities within each zone, and the level of service data providing travel cost and travel distance information among zones. The zone structure and network system, including a hierarchy of zones that was generated by using Geographic Information System (GIS) software, were provided by NCTCOG. Among the hierarchy of the zones, the smallest zone unit corresponds to Traffic Survey Zone (TSZ). The DFW metropolitan area contains a total of 5,999 TSZs. The second smallest zone unit in the hierarchy system is the Transportation Analysis Process (TAP) zone. There are a total of 919 TAPs in the study area. Our modeling processes are based on the TAP level. In the following sections, without specification “zones” refers to TAP zones.

2.2 TRIP GENERATION

Trip generation models translate demographic data (e.g. household size, income etc.) into the number of person-trips produced from or attracted to each zone. The traditional trip generation step includes two components: a trip production model and a trip attraction model. In this study, the trip generation step includes only the trip production model, which estimates the number of trips produced by each TAP in the region. Trip attractions are estimated by a disaggregate attraction-end choice model, which is discussed in the trip distribution section.

The trip generation outputs serve as inputs to subsequent modeling steps. Errors made in the trip generation phase, therefore, are likely to be magnified in the forecasting process. Therefore, improving trip generation modeling techniques is an important issue in travel demand modeling. In this study, ordered response choice models are used to estimate household trip productions. Although linear regression and cross-classification models are most commonly used for trip production, they suffer from some limitations. First, the number of trips produced by a household assumes only non-negative ordinal values (0, 1, 2, 3, etc.). The linear regression approach is not suited for the modeling of such a discrete-level dependent variable since the error assumption maintained in the approach is applicable only for a continuous dependent variable. Second, when used in forecasting, the linear models may predict negative trips or overestimate trips. The number of trips is not bounded in linear regression. Third, even if the predicted trips are in the acceptable range, the linear regression and cross-classification models do not provide the discrete probability distribution for the number of trips. Only the expected number of trips of each household is predicted. The ordered response choice model has none of these deficiencies. It captures discrete and ordinal nature of number of trips, and guarantees non-negative predicted trip rates. It also can provide the probability distribution of trip rate made by each household.

Another advantage of using the ordered response choice model is that a continuous accessibility measure can be included as an explanatory variable in the model. Several formulations for accessibility have been developed in the literature (Agyemang-Duah and Hall 1997; Fotheringham 1983), although these measures are seldom used in practice. In this study, an accessibility measure that accommodates a composite impedance measure instead of a simple distance measure is used. The inclusion of such a composite impedance term enables an assessment of the effect of level of service changes (such as travel time improvement and cost improvement) on the number of trips produced by a household. This is important for the evaluation of TCMs.

In the trip generation model developed in the current study, person trips are stratified by the three trip purposes used by NCTCOG. These three trip purposes are: home-based work, home-based non-work, and non-home-based trips. Furthermore, home-based non-work trips are classified into more disaggregate trip purposes. Segmented ordered response choice models are developed to identify and interpret similarities and dissimilarities of the effects of explanatory variables on trip rates for different trip purposes.

2.2.1 Ordered Response Choice Model

The ordered response choice model assumes that the latent propensity to make trips is a linear function of household attributes. This latent trip-making propensity, T_i^*, for household *i*, is written as:

$$T_i^* = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \varepsilon_i = X_i' \beta + \varepsilon_i \tag{2-1}$$

where the β's are coefficients to be estimated and the *x*'s are household attributes. ε_i is the error term, which is assumed to be normally distributed with a mean of zero and a variance of one for the ordered probit model. The error term reflects the randomly varying tastes of households due to unobserved attributes. The form of the latent propensity function is similar to the standard linear regression form.

The model assumes a series of thresholds (μ's) with the order of $\mu_m < \mu_{m+1} < \mu_{m+2}$. These thresholds correspond to 0, 1, 2, ..., *M* trips, where *M* is the maximum possible trip rate of a household. The number of household trips depends on the position of the latent propensity for a household relative to the thresholds representing different trip rates. If $T_i^* < \mu_0$ (μ_0 is set to be 0 in the model), the household *i* will make no trips. If $\mu_{m-1} < T_i^* < \mu_m$, the household will make *m* trips. To obtain a probabilistic estimation of household trip rates, the error term has to be taken into account. With the normal distribution assumption of the error term, the probability of household *i* making *m* trips is:

$$\Pr(T_i = m) = \Pr(\mu_{m-1} < X_i'\beta + \varepsilon_i < \mu_m) = \Pr(\mu_{m-1} - X_i'\beta < \varepsilon_i < \mu_m - X_i'\beta)$$

$$= \Phi(\mu_m - X_i'\beta) - \Phi(\mu_{m-1} - X_i'\beta) \tag{2-2}$$

Combining the probabilistic information of all trip rates in the feasible choice set, the expected trip rate for household i can be written as:

$$E(T_i) = M * [1 - \Phi(\mu_M - X_i'\beta)] + \sum_{m=1}^{M-1} m * [\Phi(\mu_m - X_i'\beta) - \Phi(\mu_{m-1} - X_i'\beta)] \tag{2-3}$$

2.2.2 Data Source and Sample Description

The 1996 household activity survey data and person demographic data provided by NCTCOG are used for this study. The household trip file is generated from the 1996 activity survey data. The trip file contains information about each trip made by each individual in the household. These person trips are aggregated to obtain household trips by purpose. The household demographic data are derived from person demographic data, including information on income, household size, age structure, and education level of individuals in the household. After data cleaning, there are 3,482 household records for estimation of the trip production models.

The frequencies of household trip rates by trip purposes are shown in Table 2-1. Surprisingly, an extremely high percentage of households (1,119, 32.1%) made no home-based work trips during the survey day. Another 1,115 households (32%) made 2 home-based work trips, and only 20 households (0.6%) made more than 6 home-based work trips. The number of home-based non-work trips shows larger variation, ranging from 0 to 35 trips. Some 654 households (18.8%) did not make any home-based non-work trips and 32 households (0.9%) made more than 20 home-based non-work trips, during the survey day. The sample data indicate that a large fraction of households made an even number of home-based trips, while for non-home-based trips, no such difference was found.

Based on the sample distribution, the maximum number of trips for estimation of the ordered response models is chosen to be 6 for home-based work trips, 20 for home-based non-work trips, and 14 for non-home-based trips. Table 2-1 also includes the frequency distributions for disaggregate home-based non-work trip purposes. The maximum trip rate for the disaggregate trip purposes is set to be 4.

Table 2-1. Frequencies of household trip rates by purpose

Trip Purpose	Number of trips								
	0	1	2	3	4	5	6	7	≥ 8
Home-based non-work trips	654 18.8%	315 9.0%	711 20.4%	222 6.4%	428 12.3%	139 4.0%	249 7.2%	122 3.5%	175 5.0%
Non-home-based trips	1,163 33.4%	428 12.3%	459 13.2%	312 9.0%	306 8.8%	184 5.3%	180 5.2%	119 3.4%	77 2.2%
Home-based work trips	1,119 32.1%	530 15.2%	1,115 32.0%	233 6.7%	350 10.1%	62 1.8%	53 1.5%	8 0.2%	8 0.2%
Home-based community trips	3,191 91.6%	84 2.4%	127 3.6%	14 0.4%	38 1.1%				
Home-based grocery shopping trips	2,583 74.2%	386 11.1%	375 10.8%	54 1.6%	61 1.8%				
Home-based non-grocery shopping trips	3,233 92.8%	125 3.6%	87 2.5%	14 0.4%	19 0.5%				
Home-based personal business trips	2,452 70.4%	471 13.5%	361 10.4%	64 1.8%	87 2.5%				
Home-based recreational trips	2,551 73.3%	297 8.5%	338 9.7%	51 1.5%	137 3.9%				
Home-based social trips	3,036 87.2%	172 4.9%	190 5.5%	32 0.9%	32 0.9%				

2.2.3 Empirical Results

2.2.3.1 Base Models

NCTCOG currently uses cross-classification models to estimate trip productions. The cross-classification of trip productions is based on six household sizes (household size 1 to 6+) and four income quartiles. The cross-classification form was selected by NCTCOG because this form makes it easier to aggregate household trip productions to zonal level productions with a joint distribution of household size and income quartile. In our ordered response models, the latent propensity is assumed to be a function of six household sizes and four income quartiles as a base model. With interactions between household size and income, a total of 24 explanatory variables are included. The 25th, 50th, and 75th percentile income values from 1996 U.S. census data are used to determine to which income quartile a household belongs. These threshold values are $18,640; $36,306; and $64,031.

Table 2-2 provides the estimated coefficients and thresholds for the ordered response model for each trip purpose. The model results indicate that, for all three trip purposes, as household size increases, the number of trips made by that household increases. The model results also show that higher income households are likely to make more home-based work trips and non-home-based trips than lower income households. For home-based non-work trips, no significant differences in trip rates were found among households in different income quartiles. Table 2-3 through Table 2-8 show the estimated trip rates for each trip purpose ignoring the stochastic nature of the propensity function. The trip rates shown in Table 2-6 through Table 2-8 are expected trip rates that recognize the probabilistic nature of the propensity function. These trip rates are computed based on Equation 2-3.

2.2.3.2 Models without Interaction Effects

The base models include a total of 24 explanatory variables. Each household size dummy variable interacts with each income quartile dummy variable. The estimation results show that most of the explanatory variables are significantly different from zero. However, we suspected that the interaction effects, which demonstrate the marginal effects of higher income quartile (or larger household size) on trip rates, might not be significantly different among different household sizes (or different income quartiles). So, we developed models without interaction terms for all three trip purposes. These models include a total of 10 explanatory variables instead of 24 variables. The coefficients of explanatory variables that are statistically significant in their effects on household trip rates are shown in Table 2-9.

The model results for home-based work trips and non-home-based trips confirm that higher income households and larger size households are likely to make more trips. For home-based work trips, the model results also show that for larger size households (household size ≥ 4), the marginal effect of household size is not significantly different from zero. This may be because, in reality, most large size households include two adults and several kids. The marginal effect of one more kid may not significantly affect the number of home-based work trips that is usually made by employed household members.

It is a little surprising in Table 2-9 that income has no significant impact on the number of home-based non-work trips. Many previous studies have found that income significantly affects home-based shopping trips or home-based recreational trips. One explanation would be that, because the travel cost is not considerably high, low-income households may travel as much as high-income households do. Another possible reason could be that the home-based non-work trip category is too broad for capturing the effect of income. For example, the effect of income on home-based school trips may be canceled out by the effect of income on home-based recreational trips. Later, we developed models for more disaggregate home-based non-work trip purposes and found that income does have significant effects for some of these trip purposes.

The likelihood ratio test was used to compare the base models to the models without interaction effects of income and household size. The likelihood ratio test statistic is:

$$-2\times(LL_R - LL_U) \tag{2-4}$$

where LL_R is the log-likelihood of the restricted model, and LL_U is the log-likelihood of the unrestricted model. This test statistic is chi-squared distributed. The null hypothesis, generally speaking, is that the restricted model is not significantly different from the unrestricted model. If the test statistic is greater than $\chi^2_{(d,\ 1-\alpha)}$, where d is the degrees of freedom and α is the significance level, then the null hypothesis is rejected at the significance level of α and the unrestricted model is preferred. Otherwise, the null hypothesis cannot be rejected and the restricted model is preferred.

In these tests, the models without interaction effects are restricted models and the base models are unrestricted models. The test results are shown in Table 2-10. The test statistics show that the base models are not significantly different from the models without interaction effects at a significance level of 2%. Therefore, in terms of model efficiency, the models without household size and income quartile interactions are preferred.

2.2.3.3 Models with Additional Demographic Variables

Household demographic variables have been shown to be important determinants of the number of trips made by a household (Agyemang-Duah et al, 1995). Consequently, the household size and income quartile may not be the only variables that affect household trip making behavior. Other explanatory variables that could be included in the model of trip productions are: household member's gender, age structure, and employment status. Including these variables may significantly improve the model in terms of effectiveness and efficiency. For example, including number of workers in the model for home-based work trips can substantially improve the accuracy of model estimation.

We tried various model specifications using numerous household demographic variables. The final models are shown in Table 2-11. All of these models are significantly better than the base model in terms of the log-likelihood ratio test.

The model results for home-based work trips imply that the number of employed household members has a significant impact on the number of home-based work trips. With this variable in the model, the household size and income quartile variables become statistically insignificant. We found another significant explanatory variable for home-based work trips to be the number of African-American household members. This result indicates that the home-based work trips of African-Americans are very different from those made by the other racial groups. The difference may be related to different job types among African-Americans and the other racial groups.

The model results for home-based non-work trips show that households with more kids (younger than 16) are likely to make more home-based non-work trips, and adults (older than 17) in general make fewer trips than kids. We found that people with higher education make more trips than those with lower education. The model also indicates that households with more workers make substantially fewer home-based non-work trips than households with fewer workers. And households with more students are likely to make more home-based non-work trips. All these findings are consistent with those of previous studies (Goulias et al. 1988; Agyemang-Duah et al. 1995).

For non-home-based trips, the significant variables include household income and the number of female household members. The results show that high income households make more non-home-based trips than low income households. It is found that unemployed non-student adults make fewer non-home-based trips and employed household members make more non-home-based trips. This may be because of time constraints for students and workers, leading to a higher level of trip chaining, and therefore, more non-home-based trips.

2.2.3.4 ***Models with Accessibility Measures***

A Hansen-type accessibility measure, which measures the accessibility from a residential zone j to shopping/recreational opportunities, is included in our models. The accessibility measure for zone j is formulated as:

$$M_j = \left[\frac{1}{L} \sum_{l=1}^{L} \left(\frac{\log R_l}{\log H_{lj}} \right) \right] \tag{2-5}$$

where R_l is the retail and service employment in zone l, L is the total number of zones in the DFW metropolitan area, and H_{lj} is the composite travel impedance between zone l and j. The computation of the composite impedance is discussed in detail in the trip distribution model section.

The Hansen accessibility measure was included in the trip generation models for home-based non-work and non-home-based trips. However, our modeling results show that it does not significantly impact the number of trips made by a household. The coefficients of the accessibility measure are not statistically significant in both models. This result suggests that the number of trips generated by a household is dependent only on household characteristics. The improvement of an area's accessibility, such as by imposing TCM strategies, and therefore improving the level-of-service and the composite impedance, does not appear to significantly increase or decrease the number of trips generated in the area.

2.2.4 Disaggregate Trip Purposes

It has been recognized that the trip purpose-specific models lead to better interpretations of travel behavior and are more efficient than models not specific to trip purpose. The use of disaggregate trip purposes will generally obtain more accurate estimation results. The three trip purposes used in this study are at a rather aggregated level. For example, the category of home-based non-work trips includes home-based shopping trips, home-based recreational trips, home-based personal business trips, etc. These trips may have different characteristics and combining them in one category may not be appropriate. Therefore, the home-based non-work trips are selected for further disaggregate trip purposes study. The home-based non-work trips are classified into six categories in the study. These six trip purposes are: home-based community trips, home-based grocery shopping trips, home-based non-grocery shopping trips, home-based personal business trips, home-based recreational trips, and home-based social trips.

A series of models were developed for the analysis. First, models were developed for each trip purpose. Second, a partially segmented model and an unsegmented model were estimated for each pair of trip purposes (15 pairs total) for comparison. The log-likelihood value for each model is shown in Table 2-12. The statistical tests show that all partially segmented models are significantly better than the unsegmented models, which means that the partially segmented models are preferred for model development. Finally, based on previous model results, a partially segmented model that includes all six trip purposes was developed. The model results are shown in Table 2-13.

The model results in Table 2-13 show that non-Caucasian racial variables have different contributions for different trip categories. For Caucasian households, almost all household demographic variables have the same impact on the home-based grocery shopping trips as on the home-based personal business trips. Only the number of household members older than 22 has slight different impacts on the grocery shopping trips and the personal business trips. These two trip purposes are the most similar trip purposes in terms of the impacts of all independent variables. This implies that it is reasonable to combine home-based grocery shopping trips with home-based personal business trips for trip generation model estimation.

The results in Table 2-13 also indicate that most of the independent variables have a similar impact on the grocery shopping trips and the non-grocery shopping trips. For Caucasian households, the variables with different coefficient values are the low-medium income dummy variable and the number of household members older than 22. Therefore, for households of parents with kids, the grocery shopping and the other shopping trip-making behavior do not differ very much. However, for households with different numbers of adults, the trip making behavior of these trip purposes differs substantially and it is not appropriate to combine these two trip purposes into one category.

2.2.5 Recommendations and Forecasting Needs

In this study, ordered response probit models were developed to estimate the household trip productions. Various model specifications were estimated and tested. The modeling and statistical test results indicate that it is beneficial in terms of model effectiveness and efficiency to include more household demographic variables in the model other than household size and income. These variables include the number of employed household members, the number of students, and the age structure of households.

In our study, an accessibility measure was developed and was included in the ordered response probit model. However, the model results show that the accessibility measure does not have a significant impact on the number of trips. This result suggests that the number of trips made by a household depends only on the household's demographic attributes. In other words, human needs are more influential than accessibility in determining the number of trips made by a household.

To examine the similarities and dissimilarities of explanatory variables' impacts on different trip purposes, segmented ordered response probit models were developed for six disaggregate trip purposes. By comparing the coefficients for different trip purposes, we found that it is reasonable to aggregate the home-based grocery shopping trips with the home-based personal business trips.

To aggregate household trip productions to a zonal level, a joint distribution of explanatory variables for each zone is needed. Based on our final models, we need the joint distribution of income, number of workers, number of students, race and age structure. We developed a procedure to synthesize the population using sample distributions of the relevant demographic variables. The procedure was first developed for each census tract in the DFW area. Once the joint household demographic distributions are available for each census tract, we assume the same joint demographic distribution for all the TSZs within each census tract, because a TSZ is the only spatial unit such that no TSZ intersects across two census tracts. Next, the demographic distribution of each TSZ is aggregated to obtain the demographic information for each TAP.

Two data sources are used in the procedure to synthesize the population distributions at the TAP level for the DFW area. One is the Public Use Microdata Samples (PUMS). The other is census data that contain information on the marginal distributions of different demographic characteristics, such as age, gender, and income categories. The PUMS consist of records of a 5% sample of housing units in the United States, with information on the characteristics of each housing unit and the people in the household. The PUMS are the most detailed demographic information available to us. However, for confidential reasons, the data lack geographic identifying information. The sample can only be identified at the county level, while our disaggregate modeling procedures require inputs for smaller geographic units, such as TAPs and TSZs. Therefore, to obtain joint demographic distribution for each census tract, we applied an iterative proportional fitting (IPF) method with the joint demographic distribution provided by the PUMS as a starting point, with the marginal distributions provided by census data as constraints. The C programming language is used to implement this procedure. With the joint distribution of household demographics, the household trip productions can be aggregated to a zonal level and serve as inputs for the attraction-end choice modeling step.

Table 2-2. ORP model results with income & household size variables

a. HBW Trips

Variable	Coefficient	T-value	α
Constant	-0.4633	-6.1500	0
Household Size 2 & Low Income	0.5652	5.3790	0
Household Size 3 & Low Income	0.8236	6.9070	0
Household Size 4 & Low Income	0.8955	6.6940	0
Household Size 5 & Low Income	1.0415	5.9640	0
Household Size 6+ & Low Income	0.7939	3.6340	0.0003
Household Size 1 & Low-median Income	0.4186	4.2150	0
Household Size 2 & Low-median Income	0.7988	8.1990	0
Household Size 3 & Low-median Income	1.0651	8.1140	0
Household Size 4 & Low-median Income	1.2839	8.8180	0
Household Size 5 & Low-median Income	1.0244	4.6260	0
Household Size 6 & Low-median Income	1.3851	6.7890	0
Household Size 1 & High-median Income	0.8315	7.3430	0
Household Size 2 & High-median Income	1.1181	12.3960	0
Household Size 3 & High-median Income	1.4257	13.2030	0
Household Size 4 & High-median Income	1.3381	11.2910	0
Household Size 5 & High-median Income	1.2725	8.2610	0
Household Size 6 & High-median Income	1.4975	6.9210	0
Household Size 1 & High Income	0.6206	2.8660	0.0042
Household Size 2 & High Income	1.3852	14.5540	0
Household Size 3 & High Income	1.4947	13.2260	0
Household Size 4 & High Income	1.3411	12.9200	0
Household Size 5 & High Income	1.3895	9.5890	0
Household Size 6 & High Income	1.5391	7.9540	0
Threshold 1	0.4369	25.1200	0
Threshold 2	1.4042	47.5350	0
Threshold 3	1.6889	52.3290	0
Threshold 4	2.4203	52.8050	0
Threshold 5	2.7012	49.9030	0
Log-likelihood value at converge	-5,351.636		

Table 2-2. ORP model results with income & household size variables (cont.)

b. HBNW Trips

Variable	Coefficient	T-value	α
Constant	0.2712	3.3630	0.0008
Household Size 2 & Low Income	0.5633	5.0380	0
Household Size 3 & Low Income	1.1122	9.3460	0
Household Size 4 & Low Income	1.5122	12.3960	0
Household Size 5 & Low Income	1.9128	11.5650	0
Household Size 6+ & Low Income	1.3609	6.7150	0
Household Size 1 & Low-median Income	0.0063	0.0600	0.9522
Household Size 2 & Low-median Income	0.6300	6.1070	0
Household Size 3 & Low-median Income	1.3220	9.9560	0
Household Size 4 & Low-median Income	1.5337	10.9690	0
Household Size 5 & Low-median Income	1.8156	11.2410	0
Household Size 6 & Low-median Income	2.1866	11.6470	0
Household Size 1 & High-median Income	-0.0713	-0.6220	0.5342
Household Size 2 & High-median Income	0.6084	6.3460	0
Household Size 3 & High-median Income	1.0468	9.6990	0
Household Size 4 & High-median Income	1.5998	13.7190	0
Household Size 5 & High-median Income	2.2819	15.6190	0
Household Size 6 & High-median Income	2.6230	13.8320	0
Household Size 1 & High Income	0.2475	1.3120	0.1896
Household Size 2 & High Income	0.5553	5.6460	0
Household Size 3 & High Income	1.0826	9.8170	0
Household Size 4 & High Income	1.6697	16.2160	0
Household Size 5 & High Income	2.0136	15.2380	0
Household Size 6 & High Income	2.1821	12.0580	0
Threshold 1	0.3271	19.2560	0
Threshold 2	0.9552	37.8570	0
Threshold 3	1.1502	42.4100	0
Threshold 4	1.5555	50.4590	0
Threshold 5	1.7000	53.4770	0
Threshold 6	1.9898	58.3080	0
Threshold 7	2.1477	60.0730	0
Threshold 8	2.4202	61.5650	0
Threshold 9	2.5532	62.7280	0
Threshold 10	2.7697	61.8520	0
Threshold 11	2.8735	61.9090	0
Threshold 12	3.0587	60.3180	0
Threshold 13	3.1889	60.1450	0
Threshold 14	3.3401	57.8540	0
Threshold 15	3.4168	55.9610	0
Threshold 16	3.5612	54.2300	0
Threshold 17	3.6393	53.2000	0
Threshold 18	3.8111	50.6520	0
Threshold 19	3.9195	48.3430	0
Log-likelihood value at converge	-8,087.968		

Table 2-2. ORP model results with income & household size variables (cont.)

c. NHB trips

Variable	Coefficient	T-value	α
Constant	-0.1938	-2.8750	0.0040
Household Size 2 & Low Income	0.3903	3.8840	0.0001
Household Size 3 & Low Income	0.7112	6.7050	0
Household Size 4 & Low Income	0.6989	5.8110	0
Household Size 5 & Low Income	0.9422	6.2050	0
Household Size 6+ & Low Income	0.6236	2.9070	0.0036
Household Size 1 & Low-median Income	0.2645	2.8990	0.0037
Household Size 2 & Low-median Income	0.4651	5.0380	0
Household Size 3 & Low-median Income	0.8262	7.0670	0
Household Size 4 & Low-median Income	0.8946	6.7450	0
Household Size 5 & Low-median Income	0.9505	5.3830	0
Household Size 6 & Low-median Income	1.2493	6.3910	0
Household Size 1 & High-median Income	0.4679	4.5760	0
Household Size 2 & High-median Income	0.5227	6.0050	0
Household Size 3 & High-median Income	0.8562	8.4730	0
Household Size 4 & High-median Income	0.9875	9.2390	0
Household Size 5 & High-median Income	1.4014	9.2390	0
Household Size 6 & High-median Income	1.3681	7.2980	0
Household Size 1 & High Income	0.4553	2.6130	0.0090
Household Size 2 & High Income	0.7663	8.4530	0
Household Size 3 & High Income	0.9728	9.4570	0
Household Size 4 & High Income	1.3522	14.0640	0
Household Size 5 & High Income	1.4940	10.4060	0
Household Size 6 & High Income	1.0943	4.8240	0
Threshold 1	0.3419	22.3250	0
Threshold 2	0.6960	33.9750	0
Threshold 3	0.9481	40.6950	0
Threshold 4	1.2284	46.5650	0
Threshold 5	1.4285	49.9470	0
Threshold 6	1.6650	52.4010	0
Threshold 7	1.8595	53.3170	0
Threshold 8	2.0194	53.2500	0
Threshold 9	2.1953	52.6060	0
Threshold 10	2.3513	49.8250	0
Threshold 11	2.5209	48.4360	0
Threshold 12	2.7036	45.4410	0
Threshold 13	2.8794	42.8370	0
Log-likelihood value at converge	-7,428.197		

Table 2-3. Predictions for HBW trips

	HH Size1	HH Size2	HH Size3	HH Size4	HH Size5	HH Size6+
Low Income	0	1	1	1	2	1
Low-Median Income	0	1	2	2	2	2
High-Median Income	1	2	2	2	2	2
High Income	1	2	2	2	2	2

Table 2-4. Predictions for HBNW trips

	HH Size1	HH Size2	HH Size3	HH Size4	HH Size5	HH Size6+
Low Income	1	2	4	6	7	5
Low-Median Income	1	2	5	6	7	8
High-Median Income	1	2	4	6	8	12
High Income	2	2	4	6	7	8

Table 2-5. Predictions for NHB trips

	HH Size1	HH Size2	HH Size3	HH Size4	HH Size5	HH Size6+
Low Income	0	1	2	3	3	3
Low-Median Income	1	1	2	3	3	4
High-Median Income	1	1	2	3	4	4
High Income	1	2	3	4	5	3

Table 2-6. Predictions for HBW trips with consideration of the error term effects

	HH Size1	HH Size2	HH Size3	HH Size4	HH Size5	HH Size6+
Low Income	0.555	1.077	1.380	1.470	1.662	1.343
Low-Median Income	0.922	1.349	1.694	2.001	1.639	2.149
High-Median Income	1.389	1.766	2.209	2.079	1.984	2.317
High Income	1.139	2.149	2.313	2.084	2.155	2.380

Table 2-7. Predictions for HBNW trips with consideration of the error term effects

	HH Size1	HH Size2	HH Size3	HH Size4	HH Size5	HH Size6+
Low Income	1.816	3.155	4.921	6.504	8.324	5.876
Low-Median Income	1.828	3.345	5.721	6.596	7.863	9.679
High-Median Income	1.679	3.283	4.686	6.883	10.167	11.942
High Income	2.348	3.133	4.814	7.193	8.814	9.657

Table 2-8. Predictions for NHB trips with consideration of the error term effects

	HH Size1	HH Size2	HH Size3	HH Size4	HH Size5	HH Size6+
Low Income	1.248	2.026	2.857	2.822	3.560	2.613
Low-Median Income	1.748	2.205	3.196	3.408	3.586	4.615
High-Median Income	2.211	2.348	3.288	3.707	5.182	5.056
High Income	2.181	3.017	3.659	4.996	5.539	4.066

Table 2-9. Models without interaction effects

Variable	Home-Based Work Trips			Home-Based Non-Work Trips			Non-Home-Based Trips		
	Coeff.	T-value	α	Coeff.	T-value	α	Coeff.	T-value	α
Constant	-0.3553	-6.61	0	0.2658	6.049	0	-0.1118	-2.259	0.0239
HHSIZE2	0.4584	8.546	0	0.5926	11.062	0	0.2023	3.987	0.0001
HHSIZE3	0.6931	10.964	0	1.1153	18.248	0	0.5050	8.717	0
HHSIZE4	0.6475	9.812	0	1.5984	25.738	0	0.6908	11.295	0
HHSIZE5	0.6894	9.289	0	2.0404	26.031	0	0.9136	11.242	0
HHSIZE6	0.6894	9.289	0	2.1307	22.112	0	0.7635	7.382	0
Low-Median Income	0.3003	5.457	0	-	-	-	0.1782	3.422	0.0006
Hi-Median Income	0.6028	11.555	0	-	-	-	0.2886	5.733	0
High Income	0.7011	12.698	0	-	-	-	0.4711	8.858	0
LL Value	-5,364.886			-8,105.578			-7,438.462		

Table 2-10. Results of likelihood ratio tests

	Home-Based Work Trips		Home-Based Non-Work Trips		Non-Home-Based Trips	
	Base Model	Model b	Base Model	Model b	Base Model	Model b
Log-likelihood function at convergence	-5,351.636	-5,364.886	-8,087.968	-8,104.467	-7,428.197	-7,438.462
Test statistic	26.5		32.99		20.53	
Degrees of freedom	16		18		15	
Significance level	0.05		0.02		0.15	

Table 2-11. Final model results

Variables	HB Work Trips			HB Non-Work Trips			Non-Home-Based Trips		
	Coeff.	**t-value**	**α**	**Coeff.**	**t-value**	**α**	**Coeff.**	**t-value**	**α**
Constant	-0.7421	-18.345	0	-0.0768	-1.423	0.1547	-0.1957	-3.213	0.0013
Household Age Structure Variables	-	-	-	-	-	-	-	-	-
Number of children (0-16 years old)	-	-	-	0.3471	16.196	0	0.1401	4.879	0
Number of adults (17-21 years old)	-	-	-	0.1228	2.934	0.003	-	-	-
Number of adults (22 and up)	-	-	-	0.1228	2.934	0.003	-0.1817	-4.558	0
Race Variables	-	-	-	-	-	-	-	-	-
Number of African-Americans	-0.1544	-9.39	0	-	-	-	-	-	-
Number of Non-Caucasians	-	-	-	-0.0845	-5.522	0	-	-	-
Education Level	-	-	-	-	-	-	-	-	-
Number of high school graduates	-	-	-	0.3973	12.735	0	0.1452	4.056	0
Number of college (& up) graduates	-	-	-	0.5515	16.736	0	0.3154	8.322	0
Others	-	-	-	-	-	-	-	-	-
Number of workers	1.1762	51.124	0	-0.2563	-10.386	0	0.2482	8.97	0
Number of students	-	-	-	0.5372	20.639	0	0.1940	7.141	0
Number of licenses	-	-	-	0.1012	2.499	0	-	-	-
Females	-	-	-	-	-	-	0.0616	2.097	0
Income Quartiles	-	-	-	-	-	-	-	-	-
Low-median income	-	-	-	-	-	-	0.1404	3.014	0.0026
High-median income	-	-	-	-	-	-	0.1404	3.014	0.0026
High income	-	-	-	-	-	-	0.2433	4.177	0
Log-likelihood at convergence	-4,598.832			-7,764.142			-7,270.178		

Table 2-12. Log-likelihood ratio for paired trip purposes

Paired Trip Purposes	Partially Segmented Model		Unsegmented Model		Log-likelihood value at constant
	Log-likelihood value	Number of variables	Log-likelihood value	Number of variables	
Grocery Shopping & Non-Grocery Shopping Trips	-4,084.837	12	-4,307.618	8	-4,397.455
Grocery Shopping & Personal Business Trips	-6,126.11	12	-6,145.766	8	-6,318.866
Grocery Shopping & Community Trips	-4,276.283	13	-4,438.123	9	-4,559.666
Grocery Shopping & Recreational Trips	-6,041.590	14	-6,076.881	9	-6,222.780
Grocery Shopping & Social Trips	-4,756.636	19	-4,859.503	9	-4,946.590
Non-Grocery Shopping & Personal Business Trips	-4,377.471	10	-4,671.538	9	-4,793.675
Non-Grocery Shopping & Community Trips	-2,446.733	7	-2,457.056	8	-2,539.039
Non-Grocery Shopping & Recreational Trips	-4,243.512	13	-4,513.719	11	-4,624.857
Non-Grocery Shopping & Social Trips	-2,966.373	11	-3,007.144	9	-3,059.998
Personal Business & Community Trips	-4,571.593	13	-4,796.896	10	-4,948.431
Personal Business & Recreational Trips	-6,320.458	13	-6,334.531	11	-6,531.744
Personal Business & Social Trips	-5,044.664	19	-5,204.632	10	-5,321.924
Community & Recreational Trips	-4,396.390	14	-4,609.902	10	-4,746.823
Community & Social Trips	-3,115.853	16	-3,148.540	11	-3,233.704
Social & Recreational Trips	-4,901.261	21	-5,045.268	10	-5,165.237

Table 2-13. Empirical results for disaggregate trip purposes

	All Trip Purposes	HB Non-Grocery Shopping	HB Personal Business	HB Community	HB Recreational	HB Social
Constant	-1.381					
High Income	-	-	-	-	0.189	-0.243
					(4.138)	(-2.883)
High-Median Income	-	-	-	-0.151	0.189	-0.287
				(-2.321)	(4.138)	(-4.561)
Low-Median Income	0.155	-0.253	-	-0.253	-	-0.287
	(5.077)	(-3.260)		(-3.378)		(-4.561)
Age 6 to 11	-	-	-	0.178	-	-0.15
				(3.825)		(-2.476)
Age 12 to 16	0.092	-	-	0.192	-	-
	(5.133)			(3.602)		
Age 16 to 21	0.092	-	-	-	-	-
	(5.133)					
Age above 22	0.195	-0.369	-0.09	-0.269	-	-
	(8.527)	(-17.812)	(-5.308)	(-7.073)		
Asian	-	-	-0.162	-0.192	-0.146	-0.208
			(-3.148)	(-2.944)	(-3.458)	(-2.938)
African-American	-0.063	-	-0.078	-	-0.08	-
	(-4.452)		(-2.833)		(-2.939)	
Hispanic Origin	-0.04	-	-0.112	-	-	-
	(-2.594)		(-3.14)			
High School	0.29	-	-	-	-0.121	-0.233
	(12.932)				(-4.257)	(-5.432)
College and Up	0.368	-	-	0.083	-	-0.265
	(15.979)			(2.089)		(-5.41)
Number of Workers	-0.201	-	-	-0.11	0.084	0.154
	(-10.715)			(-2.541)	(2.77)	(3.79)
Number of Students	-	-	-	-	0.119	0.143
					(5.017)	(3.85)

Thresholds	
Threshold 1	0.355
	(40.891)
Threshold 2	0.936
	(57.061)
Threshold 3	1.095
	(58.009)
LL at Convergence	-13,594.63
LL at Constant	-14,566.68

2.3 TRIP DISTRIBUTION

2.3.1 Model Structure

A multinomial logit model structure with non-linear parameters is adopted for the proposed disaggregate attraction-end choice model. The feasible choice alternatives are all 919 TAP zones in the area. It is impossible to estimate the model with such a large number of alternatives. However, if the error term in the utility function is assumed to be identically and independently distributed (IID) across zones, estimation with only a subset of alternatives in the choice set will lead to consistent model estimation (McFadden 1978). In our modeling procedure, we make such an IID assumption for the error term and use a subset of seven alternatives randomly drawn from the full set of TAPs. The seven attraction-end zones include one chosen attraction-end zone and six other randomly selected attraction-end zones.

The utility function of a candidate attraction zone *j*, for an individual *q* in production zone *i*, is written as:

$$U_{ijq} = V_{ijq} + \varepsilon_{ijq} = \mu' Z_{ijq} + \eta \log D_j + \varepsilon_{ijq} \tag{2-6}$$

where Z_{ijq} is a vector which contains a) travel impedance measures between zone *i* and zone *j*, b) interactions of the socio-demographic attributes of individual *q* with the travel impedance, and c) a zonal spatial measure; D_j is the number of elemented attractions in zone *j*; ε_{ijq} is a random term with an IID Gumbel distribution across alternatives and individuals; and μ and η are the coefficients that need to be estimated. The problem with the number of elemental attractions D_j is that it is not easily quantifiable. In this study D_j is captured by zonal size measures such as the number of employees and the total acreage by land use type. Specifically, D_j is written as the production of d_j and δ, where d_j is a vector of land use variables of zone *j* and δ is a vector of corresponding coefficients. With this substitution, Equation 2-6 can be written as:

$$U_{ijq} = V_{ijq} + \varepsilon_{ijq} = \mu' Z_{ijq} + \eta \log(\delta' d_j) + \varepsilon_{ijq} \tag{2-7}$$

The probability of individual *q* choosing zone *j* as an attraction-end from choice set C_i for a trip produced in zone *i* can be expressed as:

$$P_{ijq} = \frac{e^{v_{ijq}}}{\sum_{j' \in c_i} e^{v_{ij'q}}} \tag{2-8}$$

In our study, three models were estimated for HBW, HBNW, and NHB trips respectively. Next, models were developed for more disaggregate trip purpose classifications. The similarities and dissimilarities of the impacts of various variables among different trip purposes were compared and tested, which provides the recommendations of appropriate trip purpose classifications for trip production.

The models are estimated using the maximum likelihood method. The maximization of log-likelihood function is achieved using GAUSS programming language.

2.3.2 Data Sources and Sample Descriptions

The person trip file generated from 1996 activity survey data was used for model estimation. For home-based trips, the trip maker's home zone is defined as the production zone and the non-home zone is defined as the attraction zone. For non-home-based trips, the origin zone is defined as the production zone and the destination zone is defined as the attraction zone. Furthermore, each record of the person trip file was replicated seven times. For each set of the seven identical records, the chosen attraction zone is kept only for the first record; for the remaining six records, the chosen attraction zone was replaced by randomly selected attraction zones. Finally the attraction zone characteristics were appended to the file based on the attraction zone ID. The level-of-service (LOS) variables were also added to the file for each origin/destination zone pair. The final trip file for the model estimation includes three types of variables for model estimation: trip related variables such as LOS variables, trip makers' personal demographic information, and demographic attributes of attraction zones.

The final trip files include 4,561 person trips (4,561*7 = 31,927 records) for the HBW purpose; 7,354 person trips (7,354*7 = 51,478 records) for the HBNW purpose; and 4,239 person trips (4,239*7 = 29,673records) for the NHB purpose. For disaggregate trip purposes, the final samples include 906 home-based grocery shopping trips, 300 home-based non-grocery shopping trips, 1428 home-based recreational trips, 389 home-based social trips, 1,114 home-based personal business trips and 407 home-based community trips.

Table 2-14 shows the sample description for HBW, HBNW, and NHB trip purposes. The sample description shows that almost half of all trips (43.1%) were made by individuals with lower income (income < 15K). These low income individuals made 58.2% of home-based non-work trips. One can also observe that more than half of all trips (66.9%) were made by employed household members. The employed household members made 98.3% home-based work trips and 72.8% non-home-based trips. Table 2-15 shows the sample description for disaggregate trip purposes. It is notable that almost half (48.9%) of all home-based optional trips were made by low income individuals (income < 15 K). The sample distribution also shows that women made slightly more home-based optional trips (55%) than men, especially for home-based grocery shopping and other shopping trips (63.1% & 61.3%).

2.3.3 Independent Variables

2.3.3.1 Composite Impedance

A composite travel impedance is used in this project instead of a simple distance-based measure of a highway mode. The composite impedance combines multiple impedance measures including in-vehicle travel time (IVTT), out-of-vehicle travel time (OVTT), and travel cost (COST) of each available mode between zone pairs. First, the different impedance measures (IVTT, OVTT, and COST) for each mode were transformed into a single measure. In our study, the equivalent IVTT units for highway and transit mode were computed as follows:

$$Hwy = IVTT_{hwy} + \alpha_h * OVTT_{hwy} + \eta_{hp} * ParkingCost + \eta_{hf} * FuelCost$$

$$Tran = IVTT_{tran} + \alpha_t * OVTT_{tran} + \eta_t * FairCost_{tran} \tag{2-9}$$

where α's are the ratio of the parameters of OVTT and IVTT, η's are the ratios of COST parameters (highway fuel cost, parking cost, and transit fare cost) and IVTT parameter. The parameters were obtained from mode choice models that are currently used by NCTCOG. For HBW trips, $\alpha_h = \alpha_t = 1.8618$, $\eta_{hp} = 0.3916$, $\eta_{hf} = \eta_t = 0.1567$; for HBNW trips, $\alpha_h = \alpha_t = 2$, $\eta_{hp} = 1.5625$, $\eta_{hf} = \eta_t = 0.0625$; for NHB trips, $\alpha_h = \alpha_t = 2$, $\eta_{hp} = 0.5773$, and $\eta_{hf} = \eta_t = 0.3577$. Second, the equivalent IVTT impedance for all available models were combined into a composite impedance measure H that is formulated as:

$$H = (1-\theta_t) * Hwy + \theta_t * \left(\frac{Hwy}{1 + \frac{Hwy}{Tran^{\beta}}} \right) \tag{2-10}$$

where θ_t is a dummy variable that takes a value of 1 if transit is available and 0 otherwise. β is a positive parameter that indicates the relative weight placed on transit mode. Here we chose $\beta = 1.0752$ for HBW purpose and $\beta = 1.6155$ for HBNW and NHB purposes. These values are adopted from previous study by Bhat et al (1998a). In their study it is found that these two β's are significantly different from 1, which indicates that the highway mode determines attraction-end choice more than transit mode. Their study also found that log-linear form of the composite impedance performed substantially better than the linear form. Therefore, we included the composite impedance in log-linear form as an exogenous variable in the utility function.

2.3.3.2 Interaction of Individual Demographic Variables with Impedance

Previous studies (Bhat 1998a) have found interaction effects of individual characteristics with impedance measures and zonal attributes in attraction-end choice model. Some researchers use the IID distributed error term ε to capture variations across individuals and alternative zones. However, in fact some variations across individuals are systematic and can be captured by socio-demographic variables of individuals. For example, it has been observed that women work closer to home than men do. This phenomenon can be captured in the model by introducing an interaction term of gender and impedance. Therefore, we introduce interaction terms of composite impedance with gender, income segments, and employment status in our modeling efforts.

2.3.3.3 Zonal Size Measures

As discussed in the model structure section, the zonal size measures represent proxy measures of the number of elemental destinations within a zone. In our study, we introduce total zonal employment as a size measure for the HBW purpose. Zonal retail plus service employment is used for the HBNW and NHB purposes. In addition, other size measures such as the zonal retail area and office area are also included in the models.

2.3.3.4 Zonal Spatial Structure Measure

The zonal spatial structure measure is used to accommodate the position of attraction zones relative to one another. Because the attraction-ends for HBW trips are work places which are related to individuals' long term decisions and therefore are not likely to be affected by the spatial structure measure of work place zones, the zonal spatial structure variable is included for HBNW and NHB purposes only and not for the HBW purpose.

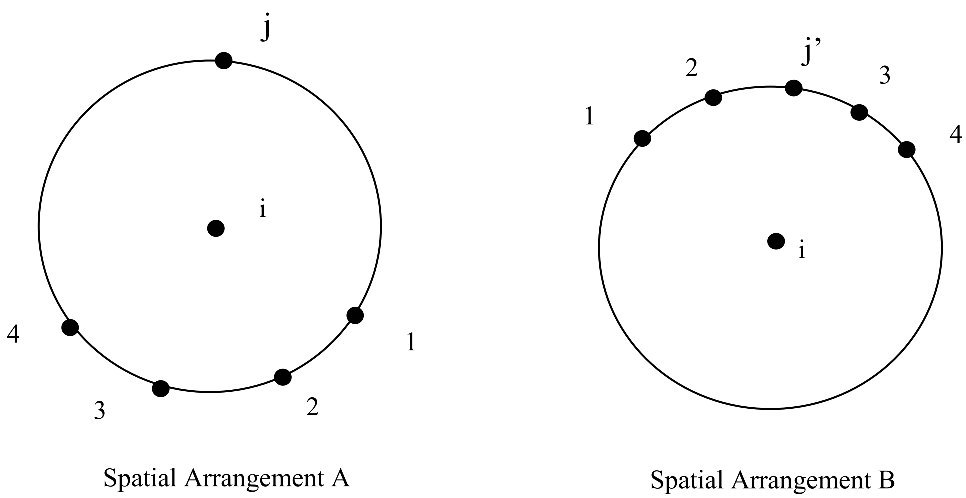


Figure 2-1. Comparison of two spatial arrangements

Figure 2-1 illustrates the motivation for including a zonal spatial structure measure in the attraction-end choice model. For two attraction alternative zone *j* and zone *j'* with identical zonal attributes and identical impedance to a production zone *i*, the conventional gravity model will estimate the identical production-attraction trip volumes among zone pairs *ij* and *ij'* even if the spatial arrangements for zone *j* and zone *j'* are substantially different. In spatial arrangement A, zone *j* is isolated from the other zones while in spatial arrangement B, zone *j'* is surrounded by some other zones. It is possible that individual's choice of zone *j* may be different for these two spatial configurations. We might conjecture that the choice probability of zone *j* may be higher than that of zone *j'* because of a "competition" effect; that is, zone *j* may occupy a unique location in the spatial cognitive perception of the individual in zone *i* because it is isolated, while there is more competition among other potential attraction zones (zone 1 through zone 5). In addition, zone *j* may be preferred in arrangement A because of higher traffic congestion in and around a group of zones (zone 1 through zone 4). On the other hand, the choice probability of zone *j'* might be higher because of an "agglomeration" effect; that is, the closely clustered zones may provide more activity opportunities, therefore zone j' might attract more trips. To illustrate the impact of the zonal spatial structure, a Hansen-type accessibility measure is introduced as the proximity of attraction alternative zone *j* to other activity opportunities. The Hansen-type accessibility measure is formulated as Equation 2-4 in the trip generation section. A zone with a large value of Hansen-type accessibility measure has more shopping/recreation opportunities in close proximity to that zone, while the lower value of Hansen-type accessibility measure indicates that the zone is spatially isolated from other shopping/recreation opportunities. Consequently, a positive coefficient of this spatial measure shows the existence of “agglomeration” effects while a negative coefficient indicates there are “competition” effects.

2.3.4 Empirical Results

The empirical results are organized into two sections. The first section presents the model results for the aggregate classification of trips into the HBW, HBNW, and NHB trip purposes. The results of NHB trip origin zone choice model are also included in this section. The second section presents the analytical results for the more disaggregate trip purposes (the six trip purposes are the same as those used in trip generation).

2.3.4.1 Attraction-end Choice for HBW, HBNW, and NHB Trips

The estimation results of attraction-end choice models for the HBW, HBNW and NHB trips are shown in Table 2-16. The model results show that the coefficient on the log of composite impedance is negative for all three trip purposes, which indicates that a larger impedance between the production zone and an attraction-end zone will make the attraction alternative less attractive. The coefficient of log of composite zonal size measure is positive, which shows that the zones with more employment and larger area will attract more trips. Among the zonal attributes, the coefficient of total zonal employment is constrained to be 1 for identification reason for HBW trips. Similarly, for HBNW and NHB trips, the coefficients of zonal retail and service employment are constrained to be 1. For HBW trips, zones with larger office area and industrial area are more likely to be selected as attraction-ends. This must be because more job opportunities are available in these zones. The coefficient of the zonal area variable indicates the number of employment that is equivalent to one square mile of zonal area in terms of zonal size representation. For example, for HBW trips, one square mile office area represents the same zonal size effect as 42.6 employees. For HBNW trips, more zonal retail and service area can significantly increase the utility of attraction-end choice while for NHB trips the model results indicate that zones with larger office, retail, and institute area will attract more NHB trips.

We found that some of the socio-demographic interactions with impedance are statistically significant. For instance, females are more sensitive to composite impedance for HBW trips and NHB trips. However, for HBNW trips, no significant difference in sensitivity to impedance was found between females and males. The model results also indicate that higher income individuals are less sensitive to impedance. For HBW trips, individuals with higher education tend to be less sensitive to impedance, which means they will travel longer distances for work than individuals with less education.

2.3.4.2 Attraction-end Choice for Disaggregate Trip Purposes

To study attraction-end choice for more disaggregate trip purposes, first a fully-segmented model was developed. Table 2-17 shows the empirical results of an unrestricted model for disaggregate trip purposes. As we expect, the coefficients of composite impedance are highly negative for all disaggregate trip purposes. The results also show that zones with larger retail and institute area are likely to attract more trips than those with smaller retail and institute area. With regards to the interactions of impedance and income for HB community and HB social trips, the results suggest that individuals with higher income are less sensitive to impedance than individuals with lower income. However, for HB recreational and HB grocery shopping trips, individuals in the lowest income group (income < 15 k) are least sensitive to impedance. A explanation is that lowest-income individuals tend to travel farther to find better deals for groceries and recreational activities.

As was done in the disaggregate trip purpose study for trip generation, different trip purposes were paired for analysis. A fully-segmented model and a restricted model were developed for each pair of trip purposes. The results show that the restricted models perform better on grocery shopping trips and non-grocery shopping trips, and recreational trips and social trips. Based on these results, a restricted model was developed for six trip purposes. Table 2-18 shows the empirical results. The log-likelihood ratio tests show that the HB recreational trips are similar to HB social trips in terms of attraction-end choice. The tests also indicate that the HB grocery shopping trips are similar to HB non-grocery shopping trips for attraction-end choice.

In summary, we have undertaken studies of disaggregate trip purposes for both trip generation and trip distribution. However, the findings from the trip generation models are not consistent with those from the trip distribution models in terms of the effect of exogenous variables. In trip generation, grocery shopping trips and personal business trips are similar, while grocery shopping trips and non-grocery shopping trips have similar attraction-end choice in trip distribution model.

2.3.5 Procedure to Apply Attraction-end Choice Model

The application of the model includes the following tasks. First, trip-makers are divided into S groups based on their socio-demographic characteristics at the zonal level. Each individual within the same group s has the same income level, age level, etc. Therefore, the interaction term of impedance and personal demographic variable is constant within the group. Second, the model is applied to get the number of trip interchanges for each group. Third, the number of trip interchanges is added across groups to obtain total trip interchanges among zones.

For an individual in group s, the utility of choosing zone j as an attraction-end for a trip from zone i can be written as:

$$V_{ijs} = \alpha_s \ln H_{ij} + \eta \log(\delta' d_j) + \mu' Z_j \tag{2-11}$$

where H_{ij} is the impedance between zone i and zone j, d_j's are zonal characteristics of the attraction-end, Z_j is spatial measure of the attraction-end, α_s is the sum of the coefficients of impedance and applicable interaction terms, and η, δ, μ are the coefficients estimated in the model.

The probability of selecting zone j as the attraction-end is given by:

$$P_{ijs} = \frac{e^{v_{ijs}}}{\sum_{j' \in c_i} e^{v_{ij's}}} = \frac{H_{ij}^{\alpha_s} (\delta' d_j)^{\eta} e^{\mu' z_j}}{\sum_{j' \in c_i} H_{ij'}^{\alpha_s} (\delta' d_{j'})^{\eta} e^{\mu' z_{j'}}} \tag{2-12}$$

where C_i is a feasible attraction-end choice set for production zone i. If O_{is} is the total number of trip productions from zone i made by individuals in group s, the total number of trips attracted to zone j can be written as:

$$T_{ijs} = O_{is} P_{ijs} = \frac{O_{is} H_{ij}^{\alpha_s} (\delta' d_j)^{\eta} e^{\mu' z_j}}{\sum_{j' \in c_i} H_{ij'}^{\alpha_s} (\delta' d_{j'})^{\eta} e^{\mu' z_{j'}}} \tag{2-13}$$

The total number of trips from zone i to zone j is the sum of trip interchanges for all socio-demographic groups. The total number of trips can be expressed as:

$$T_{ij} = \sum_s T_{ijs} = \sum_s O_{is} P_{ijs} = \sum_s \frac{O_{is} H_{ij}^{\alpha_s} (\delta' d_j)^\eta e^{\mu' z_j}}{\sum_{j' \in c_i} H_{ij'}^{\alpha_s} (\delta' d_{j'})^\eta e^{\mu' z_{j'}}} \tag{2-14}$$

The characteristics of the socio-demographic groups are shown in Table 2-19. The coefficients of the impedance for the corresponding socio-demographic groups are also shown in Table 2-19. There are a total of 12 socio-demographic groups for HBW trips, 4 socio-demographic groups for HBNW trips, 18 socio-demographic groups for the NHB attraction-end choice model, and 24 socio-demographic groups for the NHB origin zone choice model.

2.3.6 Procedure of Converting Productions/Attractions to Origins/Destinations

The disaggregate attraction-end choice models estimate the number of trip interchanges among productions and attractions. For home-based trips, it is assumed that half of home-based trip productions originate from the home zone and half of home-based trip productions' destination is the home zone. Consequently, to convert Productions/Attractions (P/A) to Origins/Destinations (O/D) for HBW and HBNW trips, half of the trips from production zone i to attraction zone j are converted to O/D trip interchange T_{ij} and another half of the trips are converted to O/D trip interchange T_{ji}. For NHB trips, the disaggregate origin-zone choice model estimates the origin zone choices of NHB trips. The attraction-ends are the same as the destination zones. Therefore, the P/A matrix directly provides the O/D matrix for NHB trips.

Table 2-14. Sample description for HBW, HBNW, and NHB trips

	HBW	HBNW	NHB	Total
TOTAL	4,561	7,354	4,239	16,154
EDUCATION				
High School	2,277 49.9%	2,651 36.0%	1,747 41.2%	6,675 41.3%
College and up	2,032 44.6%	2,383 32.4%	1,918 45.2%	6,333 39.2%
ETHNICS				
Caucasian	3,828 83.9%	6,120 83.2%	3,631 85.7%	13,579 84.1%
Asian	120 2.6%	172 2.3%	91 2.1%	383 2.4%
African-American	321 7.0%	512 7.0%	301 7.1%	1134 7.0%
Hispanics	217 4.8%	382 5.2%	133 3.1%	732 4.5%
Other race	75 1.6%	168 2.3%	83 2.0%	326 2.0%
OTHER				
Employed	4,485 98.3%	3,247 44.2%	3,082 72.8%	10,814 66.94%
Student	229 5.0%	2,279 31.0%	669 15.8%	3,177 19.7%
Female	1,998 43.8%	4,063 55.2%	2,300 54.3%	8,361 51.8%
INCOME				
Income < 15 K	1,121 24.6%	4,282 58.2%	1,553 36.7%	6,956 43.1%
Income < 25 K	977 21.4%	929 12.6%	701 16.5%	2,607 16.1%
Income < 35 K	884 19.3%	750 10.2%	652 15.4%	2,286 14.2%
Income < 50 K	810 17.8%	702 9.6%	654 15.5%	2,166 13.4%
Income < 75 K	494 10.8%	428 5.8%	435 10.3%	1,357 8.4%
Income > 75 K	275 5.9%	263 3.6%	244 5.8%	782 4.8%

Table 2-15. Sample description for disaggregate trip purposes

	HB Grocery Shopping	HB Other Shopping	HB Recreational	HB Social	HB Personal Business	HB Community	Total
TOTAL	906	300	1,428	389	1,114	407	4,544
EDUCATION							
High School	428 47.2%	122 40.7%	495 34.7%	148 38.0%	489 43.9%	151 37.1%	1,833 40.3%
College and up	344 38.0%	123 41.0%	593 41.6%	130 33.4%	465 41.7%	158 48.8%	1.813 39.9%
ETHNICS							
Caucasian	784 86.6%	256 85.4%	1.255 87.8%	334 85.9%	993 89.1%	349 85.7%	3,971 87.4%
Asian	24 2.6%	7 2.3%	21 1.5%	5 1.3%	10 0.9%	4 1.0%	71 1.6%
African-American	48 5.3%	11 3.7%	48 3.4%	23 5.9%	58 5.2%	30 7.4%	218 4.8%
Hispanic	30 3.3%	13 4.3%	64 4.5%	20 5.1%	31 2.8%	19 4.7%	177 3.9%
Other race	20 2.2%	13 4.3%	40 2.8%	7 1.8%	22 2.0%	5 1.2%	107 2.4%
OTHER							
Employed	456 50.3%	133 44.3%	731 51.2%	169 43.4%	557 50.0%	178 43.7%	2,224 48.9%
Student	92 10.2%	37 12.3%	311 21.8%	97 24.9%	130 11.7%	86 21.1%	753 16.6%
Female	572 63.1%	184 61.3%	709 49.6%	201 51.7%	601 53.9%	232 57.0%	2.499 55.0%
INCOME							
Income < 15 K	451 49.7%	145 49%	678 47.5%	229 58.8%	514 46.2%	203 49.9%	2,220 48.9%
Income < 25 K	175 19.4%	47 15.6%	207 14.5%	63 16.2%	160 14.4%	63 15.5%	715 15.7%
Income < 35 K	113 12.5%	33 11.0%	153 10.7%	33 8.5%	152 13.6%	58 14.2%	542 11.9%
Income < 50 K	94 10.4%	35 11.7%	194 13.6%	28 7.2%	152 13.6%	38 9.3%	541 11.9%
Income < 75 K	49 5.4%	27 9.0%	102 7.1%	23 5.9%	82 7.4%	32 7.9%	315 6.9%
Income > 75 K	24 2.6%	13 3.6%	94 6.6%	13 3.3%	54 4.9%	13 3.1%	211 4.6%

Table 2-16. Empirical results for HBW, HBNW, and NHB trips

Variable	HBW		HBNW		NHB	
	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.
Log of composite impedance	-3.6626	-82.391	-3.0577	-118.261	-2.9777	-83.261
Log of composite zonal size measure	0.8077	58.283	0.6363	33.64	0.7238	32.383
Total zonal employment	1	-	-	-	-	-
Zonal retail and service employment	-	-	1	-	1	
Zonal retail area	-	-	3.6219	27.804	2.6014	18.147
Zonal office area	3.753	20.757	-	-	2.6014	18.147
Zonal industrial area	1.0116	3.714	-	-	-	-
Zonal institute area	-	-	3.6219	27.804	2.6014	18.147
Spatial structure measure	-	-	-5.6883	-57.824	-6.2358	-56.483
Interactions of personal demographic variables with impedance						
Female	-0.4964	-39.116	-	-	-0.1114	-8.693
15k < Income < 25k	-	-	-	-	-	-
25k < Income < 35k	0.3005	22.909	-	-	-	-
35k < Income < 50k	0.3005	22.909	-	-	0.1135	6.366
50k < Income < 75k	0.3005	22.909	0.1007	6.21	0.3742	21.101
Income > 75k	0.3005	22.909	0.1007	6.21	0.3742	21.101
Student	-	-	-0.5127	-42.815	-	-
High school education	0.7859	25.948	-	-	0.2757	13.708
College education	1.1847	38.523	-	-	0.1461	7.041
Graduate school education	1.1847	38.523	-	-	0.1461	7.041
Number of observations	4,561		7,354		4,239	
Log-likelihood at convergence	-4,654.54		-6,429.34		-4,706.55	

Table 2-17. Full-segmented model results for six trip purposes

Variable	HB Grocery Shopping		HB Other Shopping		HB Recreational		HB Social		HB Personal Business		HB Community	
	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.
Log of composite impedance	-3.6264	-41.66	-5.5889	-31.431	-2.6546	-44.256	-2.0395	-19.7	-3.0164	-48.722	-2.5692	-21.414
Log of zonal size measure	0.5423	9.413	0.7313	7.755	0.5626	13.798	0.3903	6.191	0.6488	14.929	0.476	6.151
Zonal retail and service employment	1	-	1	-	1	-	1	-	1	-	1	-
Zonal retail area	5.1823	8.787	5.0278	7.751	3.6856	10.691	-	-	3.1731	9.828	-	-
Zonal institute area	-	-	-	-	2.9469	7.002	4.3553	7.102	-	-	3.5581	6.642
Spatial structure measure	-5.4656	-19.32	-2.5905	-5.453	-5.0994	-24.974	-4.3192	-12.967	-6.6945	-28.881	-4.3346	-11.068
Interactions of demographic variables												
Female	-	-	-	-	0.3393	14.564	0.3633	9.274	-	-	-0.7643	-16.865
Student Status	-	-	0.8894	9.829	-0.4555	-12.826	-	-	-	-	-1.7786	-21.65
Employment Status	0.4881	14.006	0.4029	7.294	0.2107	7.887	-	-	-	-	-	-
High school education	0.4794	10.354	2.5075	30.18	-0.5249	-14.109	-1.2662	-23.534	-	-	-	-
College Education	0.4794	10.354	2.5075	30.18	-0.2362	-6.129	-1.2662	-23.534	-	-	-1.2324	-26.293
Graduate School education	0.4794	10.354	2.5075	30.18	-0.2362	-6.129	-1.2662	-23.534	-	-	-1.2324	-26.293
15k < Income < 25k	-1.3862	-30.32	-	-	-0.2725	-9.128	0.9402	19.235	-	-	-	-
25k < Income < 35k	-1.0703	-25.8	-	-	-0.2725	-9.128	0.9402	19.235	-	-	1.1487	21.909
35k < Income < 50k	-1.0703	-25.8	-	-	-0.2725	-9.128	0.9402	19.235	-	-	1.1487	21.909
50k < Income < 75k	-1.0703	-25.8	-	-	-0.2725	-9.128	0.9402	19.235	-	-	1.1487	21.909
Income > 75k	-	-	-	-	0.2224	4.644	0.9402	19.235	-	-	1.1487	21.909
Asian	-	-	-	-	-3.03	-17.035	1.8351	12.391	-	-	-	-
Asian and Other Race	-	-	-	-	-	-	-	-	-2.0344	-17.406	-	-
Non-Caucasian and non-Asian	-	-	-	-	-	-	-1.8122	-27.614	-	-	1.6	27.024
Number of observations	906		300		1428		389		1114		407	
Log-likelihood at convergence	-670.855		-253.66		-1,440.713		-450.006		-1,032.36		-367.059	

Table 2-18. Restricted model results for six trip purposes

Variable	HB Grocery Shopping		HB Other Shopping		HB Recreational		HB Social		HB Personal Business		HB Community	
	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.	Coefficient	t-stat.
Log of composite impedance	-3.856	-50.682	-3.856	-50.682	*-2.4942*	*-49.69*	*-2.4942*	*-49.69*	-3.0164	-48.722	-2.5692	-21.414
Log of zonal size measure	0.5933	12.238	0.5933	12.238	*0.54*	*15.583*	*0.54*	*15.583*	0.6488	14.929	0.476	6.151
Zonal retail plus service employment	1	-	1	-	1	-	1	-	1	-	1	-
Zonal retail area	5.1039	11.694	5.1039	11.694	*3.6232*	*11.573*	*3.6232*	*11.573*	3.1731	9.828	-	-
Zonal institute area	-	-	-	-	*3.0245*	*8.244*	*3.0245*	*8.244*	-	-	3.5581	6.642
Spatial structure measure	-4.7381	-20.184	-4.7381	-20.184	*-4.9519*	*-28.1*	*-4.9519*	*-28.1*	-6.6945	-28.881	-4.3346	-11.068
Interactions of demographic variables												
Female	-	-	-	-	*0.3573*	*18.984*	*0.3573*	*18.984*	-	-	-0.7643	-16.865
Student Status	-	-	-	-	*-0.449*	*-16.25*	*-0.449*	*-16.25*	-	-	-1.7786	-21.65
Employment Status	0.3893	13.21	0.3893	13.21	-	-	-	-	-	-	-	-
High school education	-	-	-	-	*-0.6522*	*-23.66*	*-0.6522*	*-23.66*	-	-	-	-
High school and up education	0.7718	19.499	0.7718	19.499	-	-	-	-	-	-	-	-
College and up education	-	-	-	-	*-0.4852*	*-17.41*	*-0.4852*	*-17.41*	-	-	-1.2324	-26.293
15K < income < 75 K	-0.8077	-27.119	-0.8077	-27.119	-	-	-	-	-	-	-	-
income > 25 K	-	-	-	-	-	-	-	-	-	-	1.1487	21.909
income > 75 K	-	-	-	-	*0.4418*	*11.451*	*0.4418*	*11.451*	-	-	-	-
Asian	-	-	-	-	-2.9179	-16.496	1.636	11.889	-	-	-	-
Asian and Other Race	-	-	-	-	-	-	-	-	-2.0344	-17.406	-	-
Non-Caucasian and non-Asian	-	-	-	-	-	-	-1.717	-27.476	-	-	1.6	27.024
Number of observations	906		300		1,428		389		1,114		407	
Log-likelihood at convergence	-933.662				-1,899.926				-1,032.36		-367.059	

Table 2-19. Socio-demographic groups and corresponding impedance coefficients

Socio-demographic Groups	Coefficients
HBW Attraction-end Choice Model	
Female, income<25K, without high school education	-4.159
Female, income<25K,high school education	-3.3731
Female, income<25K,college & up education	-2.9743
Female, income>25K, without high school education	-3.8585
Female, income>25K, high school education	-3.0726
Female, income>25K, college & up education	-2.6738
Male, income<25K, without high school education	-3.6626
Male, income<25K, high school education	-2.8767
Male, income<25K, college & up education	-2.4779
Male, income>25K, without high school education	-3.3621
Male, income>25K, high school education	-2.5762
Male, income>25K, college & up education	-2.1774
HBNW Attraction-end Choice Model	
Income < 50 K, non-student	-3.0577
Income < 50 K, student	-3.5704
Income > 50 K, non-student	-2.957
Income > 50 K, student	-3.4697
NHB Attraction-end Choice Model	
Female, income<35K, without high school education	-3.0891
Female, income<35K, high school education	-2.8134
Female, income<35K, college & up education	-2.943
Female, 35K<income<50K,without high school education	-2.9756
Female, 35K<income<50K,high school education	-2.6999
Female, 35K<income<50K,college & up education	-2.8295
Female, income>50K, without high school education	-2.7149
Female, income>50K, high school education	-2.4392
Female, income>50K, college & up education	-2.5688
Male, income<35K, without high school education	-2.9777
Male, income<35K, high school education	-2.702
Male, income<35K, college & up education	-2.8316
Male, 35K<income<50K,without high school education	-2.8642
Male, 35K<income<50K,high school education	-2.5885
Male, 35K<income<50K,college & up education	-2.7181
Male, income>50K, without high school education	-2.6035
Male, income>50K, high school education	-2.3278
Male, income>50K, college & up education	-2.4574

Table 2-19. Socio-demographic groups and corresponding impedance coefficients (cont.)

Socio-demographic Groups	Coefficients
NHB Origin Zone Choice Model	
Female, income<15K, without high school education	-3.5245
Female, income<15K, high school & college education	-2.6668
Female, income<15K, Graduate school education	-2.8544
Female, 15K<income<35K, without high school education	-3.1456
Female, 15K<income<35K, high school & college education	-2.2879
Female, 15K<income<35K, Graduate school education	-2.4755
Female, 35K<income<75K, without high school education	-3.0429
Female, 35K<income<75K, high school & college education	-2.1852
Female, 35K<income<75K, Graduate school education	-2.3728
Female, income>75K, without high school education	-2.8923
Female, income>75K, high school & college education	-2.0346
Female, income>75K, Graduate school education	-2.2222
Male, income<15K, without high school education	-3.7757
Male, income<15K, high school & college education	-2.918
Male, income<15K, Graduate school education	-3.1056
Male, 15K<income<35K, without high school education	-3.3968
Male, 15K<income<35K, high school & college education	-2.5391
Male, 15K<income<35K, Graduate school education	-2.7267
Male, 35K<income<75K, without high school education	-3.2941
Male, 35K<income<75K, high school & college education	-2.4364
Male, 35K<income<75K, Graduate school education	-2.624
Male, income>75K, without high school education	-3.1435
Male, income>75K, high school & college education	-2.2858
Male, income>75K, Graduate school education	-2.4734

2.4 DEPARTURE TIME CHOICE MODEL

2.4.1 Introduction

As the preceding discussion indicates, there are many aspects of choice associated with individual trip-making. These choices include frequency, destination, mode, and route, and all of these components are explicitly modeled in the four-step urban transportation modeling procedure used by most planning organizations. However, time-of-day choice is a fifth and equally important dimension of choice that has received relatively little attention in the trip-based modeling approaches adopted by MPOs. The reason for this lack of attention to time-of-day may be traced back to the context in which the trip-based modeling framework was developed in the 1950s. The primary objective then was to evaluate alternative major capital improvements, so the focus was on predicting how alternative projects would affect overall daily travel demand levels and not on predicting shifts in travel within a day (Cambridge Systematics, Inc. 1994).

While evaluating capital improvements continues to remain an important objective of travel demand models, there has been a shift in emphasis in the past decade from evaluating long-term investment-based strategies to understanding travel behavior responses to shorter term, time-of-day specific, congestion management policies such as peak period pricing and peak period high occupancy vehicle use incentives. In addition, air quality modeling requires temporal resolution in the number of vehicle trips because a) the emissions factors (in grams per mile) to be applied to vehicle miles of travel (VMT) are sensitive to meteorological conditions (temperature and humidity) and vary considerably by time-of-day, b) the operating mode of trips are quite different across times of day (for example, a large proportion of trips in the morning and afternoon peak periods begin in the cold-start mode relative to other periods of the day), and c) photochemical dispersion models to determine ozone formation require mobile source emission levels of ozone precursor pollutants (i.e., oxides of Nitrogen and volatile organic compounds) by time of day (see Chatterjee et al. 1997).

The recognition of the need for temporal resolution in trip-making has led MPOs of some large metropolitan areas to apply fixed, aggregate-level, factors to apportion the predicted total daily travel to different times of the day. The use of such fixed factors in travel modeling represents an improvement over disregarding the time-of-day dimension entirely. However, it is still very simplistic and inadequate for a number of reasons. First, fixed factors implicitly assume that trip departure times are unaffected by employment-related and socio-demographic characteristics. This is a rather untenable assumption since it is very likely that employment and socio-demographic attributes are associated with constraints/preferences regarding time-of-day of participation in activities. It is particularly critical to accommodate these effects at a time when there are substantial changes in employment and socio-demographic attributes of the population which can lead to trip timing patterns in the future that are very different from those existing today. Assuming that trip timing will remain the same in the future, therefore, can lead to inappropriate policy evaluations of congestion alleviation strategies and misinformed air quality plans (see Deakin, Harvey, Skabardonis, Inc. 1993). In addition, socio-demographics vary spatially within an urban area, resulting in spatial variations in temporal travel patterns. Fixed factors are applied uniformly over the entire area, not accommodating these spatial differences. Second, applying aggregate-level factors to apportion trips to different times of the day does not accommodate departure time switching that might occur due to non-uniform (across time-of-day) variations in roadway conditions between the estimation and forecast years. Third, if time-of-day specific transportation control measures (such as congestion pricing or peak-period pricing) are implemented, the resulting temporal displacements of trips can be evaluated only by modeling level-of-service sensitivities in departure time decisions (Bhat 1998a).

2.4.2 Model Structure

2.4.2.1 Discrete Versus Continuous Model Structures

An important issue in modeling departure time choice is the representation of the dependent variable. Time is intrinsically a continuous variable, and a decision must be made whether to retain the continuous structure, or to discretize the variable for modeling purposes. There are advantages and limitations to both a continuous model structure and a discrete model structure, as is discussed next.

Time has an underlying continuous structure, and retaining this continuous-time representation is appealing for at least two reasons. First, it does not require the rather ad hoc partitioning of the day into time intervals, as a discrete method would. Second, it provides departure time of trips at a fine temporal resolution rather than in aggregate intervals. However, there are several limitations to the use of continuous-time models. They have yet to find their way into practice, whereas discrete models are commonly used by MPOs. In addition, while researchers have used continuous-time models in the past for modeling departure time of trips, most of these studies have been confined to narrow time periods of the day. Within such narrow time periods, it may be reasonable to assume that the effect of socio-demographic and employment characteristics do not change over time so that the familiar proportional hazard continuous-time model (which assumes that the covariates change the baseline hazard by a constant factor that is independent of duration; see Hensher and Mannering 1994) may be applied. However, assuming fixed effects of socio-demographics and employment characteristics is untenable when the time domain for consideration is the entire day, as is the case in the current research. For example, the effect of children on the termination of the activity duration preceding participation in recreational activity may be much more "accelerated" during the evening than in earlier times of the day because the evening is most convenient (from schedule considerations) for joint activity participation with children. Such sudden non-monotonic accelerations (or decelerations) in the effect of variables over the course of the day cannot be captured by the traditional proportional hazard or accelerated lifetime models (the accelerated lifetime model allows time-varying effects, but specifies the time-varying effects to be monotonic and smooth in the time domain). Further, level-of-service variables change during the course of the day and one must accommodate these time-varying covariates within the duration model framework. Incorporating such time-varying coefficients and time-varying covariates in duration models poses an econometric challenge (specialized econometric software needs to be developed) and also presents specification and interpretational challenges (see Bhat 1999a).

Discrete choice models, on the other hand, have the advantage of being able to easily accommodate time-varying coefficients and covariates, even using commercially available software. In addition, discrete choice models are now commonly used in practice and a discrete departure time model can be relatively easily incorporated within the travel demand frameworks of MPOs, while continuous-time duration models are still used primarily for research purposes.

For the above reasons, and because of the widespread familiarity and use of discrete choice structures, a discrete choice representation of departure time choice is retained in this project. Within the context of a discrete choice formulation, the multinomial logit model structure is adopted to estimate the departure time choice.

2.4.2.2 The Multinomial Logit Model Structure

The MNL structure relies on three basic assumptions. First, the error components of the utility function are extreme-value (Gumbel) distributed. The extreme-value distribution is one of several possible distributions that may be used. But it is the most commonly used distribution because it leads to a closed-form model for the choice probabilities. The second assumption of the MNL is that the error components are identically and independently distributed (IID) across alternatives. Finally, the MNL assumes that the error components are IID across observations. These last two assumptions imply that the variances of the error terms are the same for all individuals and all alternatives, and that there is no correlation between the error terms of various alternatives or between the error terms of various individuals (Koppelman et al. 1999).

The MNL structure is appealing because it has a simple formulation. It provides the probability that the individual will choose a given alternative based on the observable portion of the utility of the alternatives. Using the MNL, the probability that a given individual chooses alternative *i* from a set of *J* alternatives is

$$P(i) = \frac{\exp(V_i)}{\sum_{j=1}^{J} \exp(V_j)} \tag{2-15}$$

2.4.3 Data Description

We developed models for four disaggregate trip purposes. These four trip purposes are social/recreational trips, shopping trips (including grocery shopping and non-grocery shopping), personal business trips, and community trips. For ease in presentation, the social/recreational trips will henceforth be referred to simply as the recreational trips, the personal business trips as the personal trips.

The final samples for analysis included 3178 observations for the home-based recreational trip category, 2056 observations for the home-based shopping trip category, 1848 observations for the home-based personal trip category, and 740 observations for the home-based community activity category. These final samples represent approximately two-thirds of the total number of trips in the corresponding trip categories in the original raw data. The primary reason for the substantial reduction was the lack of origin/destination TAP zone data for many trips (because of which LOS information for these trips could not be determined). However, the observed split of trips among the six time periods was approximately the same in the final sample as in the original raw data.

The distribution of departure times for the four trip categories in the final samples is presented in Figure 2-2. For both the recreational and shopping categories, the number of trips increases as the day progresses, while personal and shopping trips show a non-monotonic trend during the day. The temporal differences in trip-making among the categories highlight the need to separate non-work trips into more specific categories for analysis.

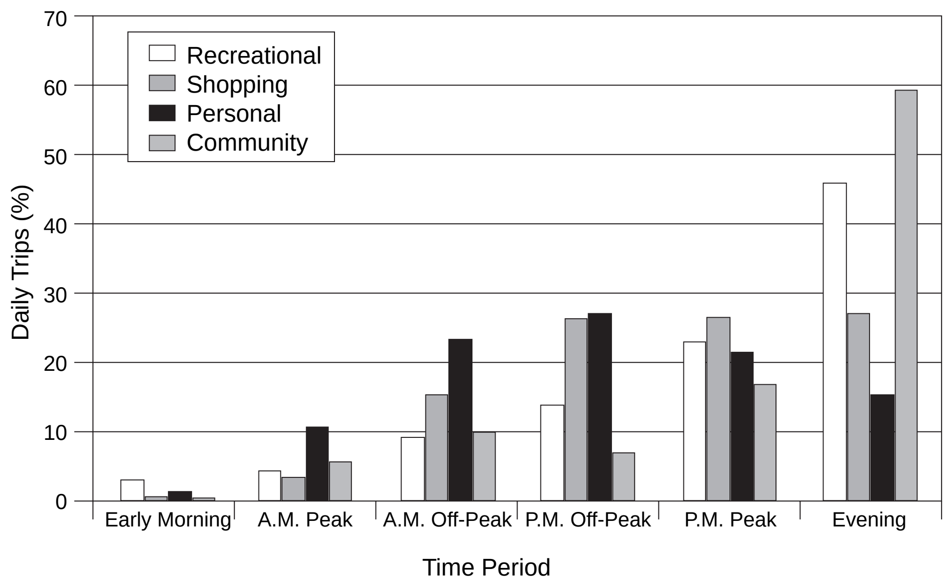


Figure 2-2. Temporal distribution of home-based recreational, shopping, personal, and community trips

The increase in trip-making as the day progresses is very noticeable for recreational trips, which have by far the greatest number taking place in the evening. For shopping trips, there is little variation in the percentage of trips among the later periods of the day (i.e., the p.m. peak, p.m. off-peak and evening periods). Personal trips experience their maximum in the p.m. off-peak, and subsequently decrease; this is quite reasonable since most businesses attracting trips classified as personal would be closed during the p.m. peak period and after. Community trips experience a minor crest in the a.m. off-peak and an overwhelming maximum in the evening. As might be expected, very few trips of any type occur in the early morning or morning peak hours. Overall, the temporal distributions of recreational and community trips are quite similar, as are the temporal distributions of shopping and personal trips. Table 2-20 shows the distribution of trips among the drive-alone, shared-ride, and walk modes. The dominant mode for recreational and community trips is shared ride, indicating joint activity participation by several family members. The dominant mode for shopping and personal trips is drive alone. This suggests less joint activity participation for shopping and personal trips than for recreation and community trips. Finally, recreational trips are more likely to be pursued using the walk mode compared to shopping or personal trips. Community trips using the walk mode were excluded from the sample, as there were not enough trips in the sample to justify the inclusion of the walk mode as an independent variable in the model.

Table 2-20. Distribution of modes used for recreational, shopping, personal, and community trips

	Mode Used		
Trip Category	Drive Alone	Shared Ride	Walk
Recreational	33.7%	59.9%	6.4%
Shopping	55.3%	42.2%	2.5%
Personal	63.8%	34.0%	2.2%
Community	43.6%	56.4%	N/A

An important note must be made here about travel mode choice. For this analysis, mode choice was considered as being exogenous to departure time choice. This decision is based on the observation that almost all non-work trips are pursued using a personal automobile (see Table 2-20). The distinction between drive alone and shared-ride modes is likely to be a reflection of how many individuals choose to participate jointly in the activity, not a conscious decision of which travel mode to use for the trip. It was assumed that the decision to engage in an activity with other individuals is made prior to the decision of what time to travel, and therefore it is assumed that travel mode choice between the drive alone and shared-ride modes is pre-determined. Also, there is little variation in walk mode characteristics across different times of the day and hence it is quite reasonable to consider the choice of the walk mode to be exogenous to departure time choice. That is, individuals are likely to first make a decision to walk or not for a recreational/shopping activity, and then determine the time of day to pursue the activity.

2.4.4 Empirical Results

The current practice in many MPOs of applying fixed factors to apportion daily travel to various time periods is equivalent, in this analysis framework, to a model specification with only constants. For each trip purpose, this restrictive model can be statistically tested against the model proposed in this project using a standard likelihood ratio test.

Table 2-21 presents the log-likelihood value at market shares (corresponding to the application of fixed factors) and the log-likelihood value at convergence for the best MNL specification. For each trip purpose category, the table also computes the likelihood ratio test value for testing the restrictive "fixed factor" model with the more general model proposed in this project. As can be clearly observed, the likelihood ratio test values far exceed the appropriate chi-squared values at any reasonable level of significance. Thus, the tests strongly reject the fixed factor models in favor of the models proposed in this project.

Table 2-21. Summary of models

	Recreational	Shopping	Personal	Community
Number of Observations	3,178	2,056	1,848	740
Log Likelihoods				
Market Shares	-4,576.52	-3,082.98	-2,988.56	-905.69
Convergence	-4,056.22	-2,767.12	-2,671.45	-715.45
Likelihood ratio test value	1,040.60	631.72	634.23	370.48
Degrees of freedom	19	17	18	19
χ^2 value at 99% confidence	36.19	33.41	34.81	36.19

Table 2-22 shows the empirical results for four trip categories. For all alternative-specific variables (i.e., for all variables except the total time variable), the evening time period is the base. In instances where only a few time periods appear for a variable, all of the excluded time periods, including the evening period, form the base. One additional point must be made about the model specification for community trips; because there were so few trips observed in the early morning time period, no alternative-specific variables were estimated for that period (it is included in the base for all variables).

Table 2-22. Estimated coefficients for departure time choice models

	Recreational		Shopping		Personal		Community	
Variables	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Constant								
Morning	-3.673	-22.18	-5.216	-13.79	-3.338	-11.38	-4.700	-9.36
a.m. peak	-4.295	-15.41	-2.757	-8.75	-3.096	-10.36	-3.995	-7.85
a.m. off-peak	-3.002	-14.59	-0.991	-4.08	-0.863	-3.89	-2.391	-7.50
p.m. off-peak	-2.227	-11.66	-0.260	-1.11	-0.252	-1.18	-4.393	-8.04
p.m. peak	-1.051	-10.57	-0.402	-2.57	0.201	0.74	-2.923	-9.32
Female								
Morning	-	-	-	-	-1.014	-2.30	-	-
Age								
Morning	-	-	-	-	0.033	7.31	-	-
a.m. peak	0.020	6.83	-	-	0.033	7.31	-	-
a.m. off-peak	0.020	6.83	0.027	6.63	0.033	7.31	-	-
p.m. off-peak	0.020	6.83	0.027	6.63	0.033	7.31	0.028	3.20
p.m. peak	-	-	0.020	5.80	0.016	2.86	-	-
Ethnicity								
Mixed race								
a.m. off-peak, p.m. off-peak	0.918	3.84	-	-	-	-	-	-
Black								
p.m. off-peak, p.m. peak	-	-	-	-	-0.510	-2.34	-	-
Asian								
p.m. peak	-	-	-	-	-	-	2.767	2.41
Other								
a.m. peak, a.m. off-peak	-	-	-	-	-0.662	-2.17	-	-
p.m. off-peak	-	-	-	-	-0.662	-2.17	1.606	2.49
Income (thousands)								
a.m. peak, a.m. off-peak, p.m. off-peak	-	-	-0.055	-2.56	-0.042	-2.84	-	-

Table 2-22. Estimated coefficients for departure time choice models (cont.)

Variables	Recreational		Shopping		Personal		Community	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Presence of children age 5 or under								
Morning	0.988	8.09	-	-	-	-	-	-
a.m. peak	0.988	8.09	-	-	-	-	1.522	4.36
a.m. off-peak	0.988	8.09	-0.407	-2.41	-	-	1.522	4.36
p.m. off-peak	0.988	8.09	-0.407	-2.41	-	-	-	-
p.m. peak	-	-	-	-	-0.724	-3.57	-	-
Presence of children age 6 to 11								
morning, a.m. peak	-0.421	-2.27	0.640	2.54	-	-	-	-
Presence of children age 12 to 15								
a.m. off-peak	-	-	-	-	-0.371	-2.57	-	-
p.m. off-peak	0.384	3.85	-0.332	-1.88	-0.371	-2.57	-	-
p.m. peak	0.384	3.85	-	-	-	-	-	-
Presence of children age 6 to 15								
a.m. peak, a.m. off-peak	-	-	-	-	-	-	-0.932	-3.12
Income (thousands)								
a.m./ p.m. peak, a.m./ p.m. off-peak	-	-	-	-	-	-	0.061	2.78
Hours worked per week								
a.m. peak	-0.029	-6.43	-0.021	-3.43	-0.017	-4.90	-0.033	-3.64
a.m. off-peak	-0.049	-12.09	-0.057	-12.18	-0.042	-11.81	-0.061	-6.23
p.m. off-peak	-0.041	-12.70	-0.048	-12.71	-0.042	-11.81	-0.046	-3.26
p.m. peak	-0.016	-6.97	-0.016	-5.68	-0.017	-4.90	-0.026	-3.83
Self-employed								
a.m. peak	-	-	-	-	0.568	1.83	-	-
a.m. off-peak, p.m. off-peak	0.994	6.18	0.721	3.49	1.108	3.66	-	-
p.m. peak	-	-	-	-	0.568	1.83	1.028	2.50
Student								
a.m. peak, a.m. off-peak	-0.768	-5.24	-1.204	-4.89	-	-	-2.079	-4.48
p.m. off-peak	-0.768	-5.24	-0.907	-4.06	-	-	-	-
p.m. peak	-	-	-	-	0.392	1.92	-	-
Homemaker								
p.m. off-peak	-	-	-	-	-	-	0.997	2.49
p.m. peak	-	-	-	-	-	-	-1.082	-2.22
Retired								
p.m. off-peak, p.m. peak	-	-	-	-	-	-	0.783	2.88
a.m. / p.m. peak, a.m. / p.m. off-peak	-	-	-	-	-	-	0.061	2.78

Table 2-22. Estimated coefficients for departure time choice models (cont.)

Variables	Recreational		Shopping		Personal		Community	
	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic	Parameter	t-statistic
Home to activity								
Morning	0.848	4.03	1.251	5.35	-	-	-	-
a.m. peak	2.386	10.35	1.251	5.35	2.926	11.69	2.593	5.27
a.m. off-peak	1.495	10.50	0.577	4.51	1.165	8.65	1.593	5.39
p.m. off-peak	0.610	5.16	-	-	0.548	4.29	-	-
p.m. peak	1.323	13.71	-	-	-	-	2.323	8.43
Drive alone mode								
Morning	1.247	12.08	1.199	4.66	-	-	-	-
a.m. peak	1.247	12.08	1.199	4.66	-	-	0.934	4.17
a.m. off-peak	1.247	12.08	0.437	3.94	0.515	3.75	0.934	4.17
p.m. off-peak	1.247	12.08	0.437	3.94	0.361	2.78	0.934	4.17
p.m. peak	0.296	2.68	-	-	-	-	-	-
Total travel time	-0.022	-2.37	-	-	-	-	-	-

2.4.4.1 Individual Socio-Demographic Variables

The first individual socio-demographic variable in Table 2-22 is a female-gender dummy variable. The results indicate that gender does not appear to play a very important role in departure time choice. Only in the personal business category is it a significant variable, and even then, only for the early morning period, in which women are less likely than men to make personal trips. This may reflect the family-associated responsibilities of women in the morning.

The age variables, in general, indicate a preference of older individuals (independent of work status) to pursue non-work trips during the middle of the day, and especially during the off-peak periods. This may be a reflection of their physiological need for more time to start the day and a desire to arrive home early because of safety/security considerations among older individuals. While there is a generic trend to stay away from the early and late parts of the day for any trip purpose, there are some differences across trip purposes. For recreational trips, the coefficients suggest that older individuals avoid the early morning, p.m. peak, and late evening periods. The effect of age on departure time for shopping trips is similar, except that the p.m. peak is preferred to the a.m. peak for these trips. Age effects for personal business activities indicate that older individuals stay away from travel during the evening period, and also prefer the earlier times of the day (morning through p.m. off-peak) to the p.m. peak. Finally, age has only a marginal effect on community trips, indicating a preference of older individuals for the p.m. off-peak. For all trip purposes, the p.m. peak is one of the favored time periods for travel for older individuals (in addition to the linear effect presented in the table, non-linear spline effects of age and dummy variables for age categories were also explored; however, these non-linear effects did not dramatically improve data fit and were also difficult to interpret).

Several ethnicity variables appear to affect departure time choice. These results are rather difficult to explain, but are retained because of the rapidly changing racial composition of the Texas and U.S. populations. The variables tested (with the base race being Caucasian) included indicators for Asian, African-American, Native American, mixed race, and “other” race (this category contains a high percentage of Hispanic individuals). For recreational trips, the only ethnicity variable having a significant effect is the variable identifying if a person is of mixed race or not. The parameter on the mixed race variable indicates that individuals with such a family heritage are more likely to pursue recreational activities during the mid-day periods. Two ethnicity variables, African-American and “other,” significantly affect departure time for personal trips. The parameter on the African-American variable indicates that these individuals are less likely to choose the p.m. off-peak for personal business activities. Individuals in the "other" race category are less likely to pursue personal activities in the a.m. peak, a.m. off-peak, or p.m. off-peak periods. The effect of ethnicity on community trips shows a preference by Asians for travel in the p.m. peak, and a preference by individuals in the "other" race category for travel in the p.m. off-peak.

The final socio-demographic variable in the table is individual income. Individual income affects departure time choice for shopping and personal trips only. For both of these categories, the effect of income indicates that individuals with higher income tend to avoid trip-making during the a.m. peak and mid-day periods. This may be the result of tighter schedule constraints of high-income-earning individuals during the mid-day. As with the age variable, spline effects were tested for the income variable to see if its effect on departure time was non-linear. However, the spline representation did not improve the model significantly, and the results were difficult to interpret.

2.4.4.2 Household-Level Socio-Demographic Variables

Several household socio-demographic attributes were tested; however, the majority of the variables appearing in the final specifications are those associated with the presence and age distribution of children. The presence of young children (less than 5 years of age) in the household affects the timing of recreational, personal, and community activities in a similar manner. The results show that individuals whose households have young children are more likely to pursue recreational trips during the earlier periods of the day (early morning through p.m. off-peak) than in the p.m. peak or evening periods. This may be related to the biological needs of young children toward the end of the day. A similar result can be observed for personal trips; the p.m. peak is the least preferable time of day for these trips. For community trips, the a.m. peak and a.m. off-peak are the preferred travel periods for households with young children. Shopping activities, however, differ from the other non-work activity types with regard to the effect of young children. The presence of young children in the household suggests a lower likelihood of participation in shopping trips during the mid-day (a.m. and p.m. off-peak) hours. Perhaps this is a result of the tendency to shop alone by an adult who must remain at home in the mid-day to take care of the needs of the young children. Instead, the individual may choose to shop at other times of the day when another adult is available to care for the children.

The effect of older children on trip timing is, in most cases, opposite of the effect of young children. Individuals whose households have children between the ages of 6 and 15 tend to pursue recreational trips during the p.m. off-peak and p.m. peak. This is quite intuitive since these periods offer convenient times for joint activity participation in recreational activities with school-going children, and are also the times when children are most likely to be participating in recreational activities that may require a ride from a parent. However, members of households with older children are less likely to pursue shopping and personal trips during the mid-day. This may be because individuals in these households tend to have other commitments during the day that preclude shopping and personal trip-making. Members of households with older children also tend not to make community trips in the a.m. peak and a.m. off-peak, likely for much the same reason.

Overall, it appears that recreational, personal, and community activities are likely to be pursued jointly with children and so participation in such activities appears to be organized around the schedule availability of children (i.e., during the day for households with young children and late in the day for households with older children). However, shopping activities may be pursued alone and so these activities are scheduled in periods when young children have fewer biological needs or during times when older children are at school. For all four categories, the number of children in the household does not have any significant impact beyond that of the presence of children.

The final household level characteristic examined was household income. Interestingly, community trips are the only category for which household income, rather than individual income, affects departure time. The parameter on household income indicates that members of higher-income households are more likely to participate in community activities during the day than in the late evening or early morning time periods. This may result from members of higher-income households (although not perhaps the high-wage earners themselves) having, in general, more schedule freedom during the day and choosing to participate in community activities then. The parameter effect obviously must be applied to working members of the household as well, but its strength may be overshadowed by the effects of some of the other, work-related variables.

2.4.4.3 Individual Employment-Related Attributes

Employment-related attributes make such a substantial contribution to departure time decisions that they are considered in this analysis as a separate category of variables. The effect of the number of hours of work variable is quite consistent across all trip categories. It indicates that individuals who are employed and have a substantial work commitment are very unlikely to participate in non-work activities during the mid-day periods (i.e., the a.m. and p.m. off-peak periods). This is a reasonable result since employed individuals are typically at work during these times. These individuals are also less likely to participate in non-work activities in the peak periods (especially the a.m. peak) relative to the evening period. Overall, individuals who are employed and work many hours are likely to participate in non-work activities during the evening period because of work schedule constraints during the earlier times of the day (technically speaking, the results suggest that working individuals are as likely to participate in activities during the early morning period as in the evening period; however, this result is simply a statistical manifestation of the extremely low number of working individuals who choose the early morning period).

The parameter indicating whether an individual was self-employed, as opposed to externally (non-self) employed, also exhibits considerable consistency across categories. Self-employed individuals are more likely to participate in recreational, shopping, and personal activities during the mid-day (a.m. off-peak and p.m. off-peak) than externally employed individuals. That is, self-employed individuals are able to "sandwich" an activity from home between periods of a.m. and p.m. work because of lesser schedule rigidity. For personal trips, this effect also carries over into the a.m. and p.m. peak periods, indicating that self-employed individuals are more likely than externally employed persons to make personal trips during the peak periods. Across trip purposes, the greatest difference in the effect of the self-employed variable is for community trips. In this category, self-employment affects only the p.m. peak period, again increasing the likelihood of participation. This may be due, in part, to the temporal distribution of community activities; very few take place prior to the p.m. peak, so the self-employment effect may be statistically negligible during those periods.

Like the employment variables, the student variable is also reasonably consistent across categories. The coefficients on this variable suggest a preference by students for the p.m. peak and evening periods for participation in recreational and shopping activities, an intuitive result because students are generally free from academic obligations at these times. Students tend to prefer the p.m. peak period for personal trips, and they avoid the a.m. peak and a.m. off-peak periods for participation in community activities. The reasoning behind this is much the same as for recreational and shopping trips. An interesting difference between students and employed individuals (externally-employed or self-employed) is that students are equally likely to participate in non-work activities (with the exception of personal trips) during the p.m. peak and evening periods, while employed individuals are more likely to choose the evening period than the p.m. peak. This difference may be attributed to the increased flexibility of students during the p.m. peak, since the typical school day of a student ends earlier than the typical workday of an employee.

In addition to the above variables, there were two attributes that had significant effects for the personal business category, but were not significant in any of the other trip categories. Individuals who are primarily occupied as homemakers prefer the p.m. off-peak for personal activity participation, while avoiding the p.m. peak. These individuals probably have more schedule freedom in the early afternoon than in the p.m. peak, when children are arriving home from school. However, it is interesting that the personal business category is the only one in which this effect was exhibited.

Finally, retired persons prefer the p.m. off-peak and p.m. peak for personal activity participation. As with the homemaker variable, this variable is present only in the personal business model.

For each trip purpose, a specification was considered that also included a dummy variable corresponding to external-employment in addition to the number of hours and self-employment variables. Such a variable would add a generic effect (i.e., independent of hours of work) of being externally employed. The resulting specification showed a marginal (though statistically significant) improvement in data fit, but also led to results that were quite difficult to interpret. Therefore, the external-employment dummy variable was excluded from the model. An alternative specification that replaced the hours-of-work variable with the external-employment dummy variable also performed well and provided easy-to-interpret results, but was statistically inferior to the current specification (however, this alternative specification might offer forecasting advantages since it only requires forecasting employment status, not hours of work).

2.4.4.4 Trip-Related Attributes

The first variable indicates whether a trip was from home to an activity (as opposed to from the activity back home). The home-to-activity variable proves to be highly significant in all four models. For recreational trips, the results indicate that trips from home to an activity are likely to be pursued before the evening periods, and especially in the a.m. peak period. This is expected, since most trips originating in the a.m. peak are likely to be leaving home. The same is true, though to a lesser extent, for the a.m. off-peak period. The increased likelihood of trips from home to a recreational activity in the p.m. peak compared to the morning and p.m. off-peak periods can be attributed to the temporal "resurgence" in recreational participation during the p.m. peak period; on the other hand, a sizable fraction of home-based recreational trips in the early morning period are late return home trips from the previous evening's recreational engagement and many p.m. off-peak home-based recreational trips are return home trips after recreational participation in the a.m. periods. The parameters on the home-to-activity variable for community trips indicate a pattern similar to that for recreational trips, with an overall preference for the a.m. peak period, and a temporal resurgence in the p.m. peak. The home-to-activity variable exhibits a different effect for shopping and personal trips. For these trip types, the impact is fairly straightforward; there is a strong preference to leave for shopping and personal activities during the earlier parts of the day (morning through a.m. off-peak for shopping trips, and a.m. peak through p.m. off-peak for personal trips).

The drive-alone dummy variable effects in the table indicate a general trend, over all trip purposes, toward the use of the drive-alone mode during the earlier periods of the day and the use of non drive-alone modes (shared-ride and walking) during the later periods of the day. This is a rather intuitive result; trips during the day are more likely to be pursued alone, while the late afternoon and evening periods are times that are most convenient for joint activity participation and for walking. The differences between trip categories are slight; for personal trips, the preference is to pursue drive-alone trips during the mid-day, while for shopping trips, the morning and a.m. peak are preferred. Both recreational and community trips show a consistent preference for the daytime (p.m. off-peak and earlier), but drive-alone recreational trips are more likely to be made in the p.m. peak than in the evening. The overall preference for making drive-alone trips early in the day supports the hypothesis that trips made later in the day are more family-oriented, while trips made early in the day are individually-oriented.

Total travel time is the only LOS variable that demonstrates any significance whatsoever, and even that is only for recreational trips (a travel cost variable was also considered, as were separate representations of in-vehicle and out-of-vehicle travel time). The negative effect of travel time in the recreational trip model is consistent with a priori expectations; individuals prefer departure times that result in shorter travel times. However, the lack of any significant effect of trip travel time on departure time choice for the other purposes is interesting. Two related issues may be at work here. First, many of these activities are organized around work constraints and other household schedule considerations, and are pursued at the most convenient time within these schedule considerations. Travel time may therefore not play a substantial role in the departure time decision. Second, in the current data set the average lengths of shopping, personal, and community trips are shorter than the average length of recreational trips. In the context of the shorter lengths of these trips, there is likely to be smaller variation in travel times across time periods (in this data, the travel time variation across time periods is lower for shopping, personal, and community trips than for recreational trips). This may be manifesting itself in the form of the lack of any travel time effect on departure time decisions for shopping, personal, and community trips. Nonetheless, it is a somewhat unexpected result that travel time has such a minor effect on departure time choice.

2.4.5 Summary

This section presents models for home-based recreational, shopping, personal business, and community trip departure time choice. The departure time alternatives are represented by six temporally contiguous discrete time periods that collectively span the entire day.

The MNL model structure is used in this study. The empirical analysis in the paper uses the 1996 activity survey data collected by the North Central Texas Council of Governments (NCTCOG) in the Dallas-Fort Worth area. Several sets of variables were considered in the model specifications, including individual and household socio-demographics, employment-related attributes, and trip-related characteristics. Important overall results from the empirical analysis are as follows:

- gender does not have an important role in departure time choice,
- older individuals are most likely to participate in non-work activities during the mid-day,

- high-income-earning individuals avoid the mid-day periods for shopping and personal business,
- individuals with very young children (under 5 years of age) in their households are unlikely to pursue most activities during the p.m. peak and evening, presumably because of the increased biological needs of young children during these late times of the day,
- individuals with children below 5 years of age are unlikely to participate in shopping activities during the mid-day,
- individuals with children above 5 years of age in their households, on the other hand, are most likely to pursue recreational activities during the p.m. peak period since this is the most convenient time to jointly participate in recreational activities,
- individuals with children above 5 years of age are unlikely to pursue recreational, personal, and community activities during the mid-day,
- employed individuals and students are most likely to participate in non-work activities during the latter parts of the day,
- self-employed individuals are more likely than externally employed individuals to "sandwich" a recreational, shopping, or personal activity between the a.m. and p.m. work periods,
- trips to a non-work activity from home tend to be made before the evening period,
- trips pursued together with others or by walk are likely to be undertaken during the p.m. peak and evening periods, and
- in the current empirical context, the only level-of-service variable that has a significant impact is trip travel time and even this applies only for recreational trips.

Due to the use of a trip-based approach and the choice of the familiar MNL model structure, the current modeling effort can be incorporated relatively easily within the travel demand model system used by most MPOs for transportation planning. The results show the models proposed in this project to be an improvement over the fixed-factor approach generally used to accommodate the time-of-day dimension of travel choice.

CHAPTER THREE - EMISSIONS MODELING IMPROVEMENTS

3.1 VMT MIX MODELING FOR MOBILE SOURCE EMISSIONS FORECASTING: FORMULATION AND EMPIRICAL APPLICATION

3.1.1 Background and Significance of Work

The integration of transportation planning and air quality planning is important for mobile-source emissions estimation. The Environmental Protection Agency (EPA) requires the use of MOBILE5 for such emissions estimation for all areas except California, which uses the EMFAC7F model.

The emissions factor models (MOBILE5 and EMFAC7F) require several traffic-related inputs, including travel speeds, vehicle miles of travel (VMT), on-road operating conditions (operating mode of vehicles, environmental conditions, and existence of inspection/maintenance programs), vehicle age distribution by vehicle class in the area of analysis, and vehicle mileage accumulation rates by vehicle class. These inputs are used to calculate emissions factors (in grams per mile of vehicle travel for each pollutant) for eight different vehicle classes. The vehicle class-specific emissions factors are then applied to the VMT accumulated by each of the eight vehicle classes and are finally aggregated across all vehicle classes to obtain the total vehicular emissions.

The level of detail at which the emissions analysis is conducted varies substantially among metropolitan regions. But the EPA requires that metropolitan planning areas rated as serious or higher in nonattainment designation for ozone and carbon monoxide estimate their mobile-source emissions using network-based transportation models. The planning organizations in these areas generally conduct their emissions analysis at an individual link level. This level involves the estimation of volumes and speeds on each network link in the metropolitan area from travel demand models, followed by the computation of link-specific emissions based on a) link VMT, b) vehicle speed on the link, c) the vehicle class-specific emissions factors, and d) VMT mix fractions in the eight vehicle classes. Of these, the link VMT and link speeds are obtained directly from the network-based travel demand models. The vehicle class-specific emissions factors are obtained from the emissions-factor models based on the various inputs listed previously. The VMT fraction by vehicle class (referred to as VMT mix in the MOBILE5 model) is a supplementary traffic-related parameter that is to be provided by the analyst.

The emissions factors for each of the three pollutants, carbon monoxide (CO), volatile organic compounds (VOC), and oxides of nitrogen (NOx), vary widely among the different vehicle classes. Consequently, the emissions analysis is very sensitive to VMT mix. For example, at high temperatures, a 2.8% change in the heavy-duty gas vehicle (HDGV) mix causes about a 10% change in the CO emissions rate, and a 4.8% change in the HDGV mix leads to about a 10% shift in the VOC emissions rate (see Chatterjee et al. 1997, p. 45). It is, therefore, important to provide accurate VMT mix values at the individual link level (see *NCHRP Research Results Digest* 1998, which identifies improvement in VMT mix modeling as an area of research that would be particularly beneficial for emissions modeling). The purpose of the current report is to propose and implement a methodology for obtaining improved link-specific VMT mix values compared to those obtained from existent methods. Specifically, the research is developing a fractional split model that predicts the VMT mix on links as a function of the functional roadway classification of the link, the physical attributes of the link, the operating conditions on the link, and the attributes of the traffic analysis zone in which the link lies. In the next section, we summarize the state of the art in VMT mix modeling and discuss the improvements that this research attempted to make.

3.1.2 State of the Art/Practice in VMT Mix Determination

The emission-factor models require the total VMT accumulated on a road to be by split across eight vehicle classes. The vehicle classes are based on the size and weight of vehicles as well as on the type of fuel used. The eight vehicle classes are light-duty gasoline vehicle (LDGV); light-duty gasoline truck, type 1 (LDGT1); light-duty gasoline truck, type 2 (LDGV2); heavy-duty gasoline vehicle (HDGV); light-duty diesel vehicle (LDDV); light-duty diesel truck (LDDT); heavy-duty diesel vehicle (HDDV); and motorcycle (MC).

The current practice in many metropolitan areas is to accept the aggregate default VMT mix computed by MOBILE5 and to apply this mix to all network links. The default VMT mix is based on national data reflecting the proportion of travel by each vehicle type in urban areas.

Another approach adopted by some metropolitan agencies is using 24-hour local vehicle-classification counts (rather than MOBILE5 default values) to determine VMT mix. The agencies then apply various factors to convert the vehicle types in traffic counts to the eight MOBILE5 vehicle classes. EPA recommends that local agencies adopt this approach because the MOBILE5 default values may not be reflective of the local traffic vehicle mix. In this local vehicle count-based approach, the VMT mix is typically stratified by the functional classification of roadways to accommodate variations across roadway classes. However, since most counts are conducted only on higher roadway classes (such as interstates and major arterials), there is inadequate information for agencies to comprehensively capture variations in VMT mix by roadway class. Values of VMT mix obtained for the higher roadway classes are applied (sometimes after ad hoc adjustments) to the lower roadway classes (such as minor arterials, collectors, and local roads).

A problem with the state of the art/practice discussed above for VMT mix determination is that it applies aggregate-level values across links in the road network in a region. In an analysis of VMT mix from 477 different count sites in the U.S., Chatterjee et al. (1997) found substantial variation in VMT mix across the sites, emphasizing the need for local determination of VMT mix values (rather than using MOBILE default values). The same study also indicates substantial variation in VMT mix even after controlling for roadway class at any given site, underscoring the need to consider explanatory factors other than roadway class in local VMT mix analyses.

The discussion above motivates the research documented in this report. Specifically, we formulate and estimate a fractional split model that determines the VMT mix ratio as a function of several informative variables, which include the physical attributes of links (such as number of lanes and whether the link is a divided road or not), the operating characteristics of links (such as link speed), aggregate area-type characterizations of the traffic survey zone in which the link lies (such as urban, suburban, and rural), and the land-use attributes of the zone (such as retail acreage and manufacturing/warehouse acreage in the zone). Such a model will facilitate accurate VMT mix computations at a fine level of geographic resolution.

3.1.3 Fractional Split Model Structure

3.1.3.1 General Background

Fractional response-dependent variables arise naturally in many transportation analyses and other analytical contexts. Examples of such variables include the proportion of freight tonnage for a commodity group moving from an origin to each of several destinations, the proportion of intercity trips made by each of several travel modes, time spent by an individual in one of several activity types (such as shopping, social-recreational, or personal business), and (as in the present analysis) the fraction of VMT accrued by each vehicle class. A characteristic of all these analyses is that the variable of interest is in the form of fractions. The sum of the fractions across all categories of the variable is equal to one, and each fraction is bounded between zero and one. In addition, one or more of the fractions may take the boundary values of zero or one. In the following discussion, we present the fractional split model structure in the context of VMT mix analysis.

Mathematically, let y_{qi} be the fraction of VMT accrued by vehicle type *i* (*i*=1,2,...,*I*) on link *q*. Let this fraction be a function of a vector x_q of relevant explanatory variables associated with attributes of the link and the traffic analysis zone in which the link lies. A common approach to analyzing fractional dependent variables is to model the log-odds ratio as a linear function (for example, see Bhat and Misra 1999):

$$E\left(\log \frac{y_{qi}}{y_{q1}}\right) = \beta_i' x_q, i \neq 1, \tag{3-1}$$

where β_i is a parameter vector to be estimated for each *i* (except for a base category which needs to be normalized to zero for identification; in the above equation, the first category is arbitrarily assigned to be the base category; *i.e.*, $\beta_i = 0$, where 0 is the null vector of the appropriate size). y_{q1} is the VMT fraction accrued by the first vehicle type. If some parameters in the β_i vectors are equal across categories, such restrictions can be imposed by jointly estimating all β_i vectors after appropriate data structuring (see Bhat and Misra 1999).

The specification in Equation 3-1 is attractive since the transformed dependent variable in the regression is unbounded and can take values anywhere on the real line as y_{q1} varies between

0 and 1. Thus, a linear regression is appropriate. However, as pointed out by Papke and Wooldridge (1996), the specification has at least two major problems. First, the dependent variable is undefined when the fraction of VMT in a vehicle class is 0/1. If the numbers of vehicle class observation combinations for which the boundary conditions prevail is small, arbitrary small adjustments may be made prior to the computation of the log-odds ratio without significantly affecting the estimated parameters. However, if there are several vehicle class-observation combinations for which the boundary conditions prevail, the adjustments can have a substantial impact on estimation. In our analysis, the fractional VMT for some vehicle types (such as buses and trucks) is 0 for a large percentage of observations (i.e., links) for which vehicle classification counts are available.[1]

A second problem with the specification in Equation 3-1 is that, even if the econometric specification in Equation 3-1 is appropriate and well defined, one cannot obtain $E\left(\mathrm{y}_{\mathrm{qi}} \mid x_q\right)$ (which is of primary interest for VMT fraction forecasting) without making additional assumptions about the distribution of the residuals, $u_{qi} = \log\left[\mathrm{y}_{\mathrm{qi}} / y_{q1}\right] - \beta_i' x_q \left(i = 2,3,...,I\right)$. If a distribution is assumed or estimated, then $E\left(\mathrm{y}_{\mathrm{qi}} \mid x_q\right)$ may be computed by first obtaining the conditional expected value (on residuals) for each fraction and then unconditioning out the residuals by integrating over the distribution of the assumed or estimated distribution for the residuals (see Bhat 2000) for an application of this method). However, this approach is either nonrobust (if an incorrect parametric distribution is assumed) or cumbersome (if a nonparametric distribution for the residuals is estimated). Also, the integration in this method will involve as many dimensions as there are vehicle types, and this can become tedious.

3.1.3.2 Quasi-Likelihood Estimation

The model we propose here for VMT mix modeling is a polychotomous extension of the binary fractional split model proposed by Papke and Wooldridge (1996). The approach does not need any ad hoc adjustment for boundary values of the dependent variable fractions, and it directly specifies a model for $E\left(\mathrm{y}_{\mathrm{qi}} \mid x_q\right)$. At the same time, the approach is easy to implement

[1]If the dependent variable represents proportions from a fixed number of groups with known group sizes, suitable adjustments have been proposed in the econometric literature (see Maddala 1983, p. 30). However, the corresponding Berkson's minimum chi-squared estimation method is not applicable when the fractions arise naturally in analysis settings (such as the current VMT mix setting) rather than as a result of the discrete grouping of disaggregate observations.

and is robust, since we make no assumptions about the distribution of y_{q1} conditional on x_q. The focus is on consistent estimation of the parameters appearing in the conditional mean specification $E(y_{qi} \mid x_q)$ and on consistent, asymptotically robust estimation of the standard errors of the conditional mean parameters.

Consider the following econometric specification:

$$E(y_{qi} \mid x_q) = G_i(\beta, x_q),\ 0 < G_i(.) < 1,\ \sum_{j=1}^{I} G_j(.) = 1, \text{ where } \beta = (\beta_2', \beta_3', ..., \beta_I')'. \tag{3-2}$$

$G_i(.)$ (*i*=1,2,...,*I*) in the above equation is a predetermined function, and the properties specified for it above ensure that the predicted fractional VMT in each vehicle class for any link will lie in the interval (0,1) and will sum to 1 across vehicle classes. The econometric model in Equation 3-2 is well defined even if y_{qi} takes on the value of 0 or 1 with positive probability. The reader will note that the specification above does not make any assumption about the true underlying conditional distribution of y_{qi} given x_q. This distribution is considered unknown and can have any underlying structure.

The β parameter vector in the conditional mean model of Equation 3-2 is estimated by maximizing a likelihood function associated with a family of probability distributions that does not necessarily contain the true unknown distribution. The label "quasi-likelihood estimation" is used for such estimations (see Gourieroux et al. 1984). Specifically, we use the multinomial log-likelihood function in the quasi-estimation:

$$L_q(\beta) = \sum_{i=1}^{I} y_{qi} \log G_i(\beta, x_q). \tag{3-3}$$

The multinomial quasi-likelihood estimator used above belongs more generally to the linear exponential family (LEF). Gourieroux et al. (1984) prove the strong consistency and asymptotic normality of the parameter estimator of the conditional mean (i.e., the elements of the β vector) obtained by maximizing $\sum_q L_q(\beta)$, as long as (and if and only if) $L_q(\beta)$ belongs to the LEF (see also Wooldridge 1991). This is a very strong result, since it is based only on the correct specification of the conditional mean function of Equation 3-2. The result holds irrespective of the true distribution of y_{qi} conditional on x_q. Of course, if we are able to correctly specify this true distribution, we can maximize the true likelihood function in order to obtain an estimator more efficient than the quasi-likelihood estimator used here. However, the disadvantage of this alternative approach (compared to the quasi-approach) is that the resulting "true-likelihood" estimator is inconsistent under an incorrect assumption for the true distribution.

Within the family of LEF-based quasi-likelihood estimators, asymptotic efficiency can be achieved if the functional form of the true conditional second-order moment (i.e., variance) of y_{qi} given x_q is known. This case is unlikely. We prefer to base our inference only on the conditional mean specification of Equation 3-2 and to propose consistent and asymptotically robust inferences for the conditional mean parameter vector β. As indicated by Papke and Wooldridge (1996), this can be achieved by computing the asymptotic variance-covariance matrix of β as $H^{-1}\Delta H^{-1}$, where H is the Hessian and Δ is the cross-product matrix of the gradients (H and Δ are evaluated at the estimated parameter values).

A final model structure issue concerns the specification of the functional form for G_i in the conditional mean specification of Equation 3-2. We use a multinomial logit functional form for G_i since this structure is easy to program and implement. In this structure, we write:

$$G_i\left(\beta, x_q\right) = \frac{e^{\beta_i' x_q}}{\sum_{j=1}^{I} e^{\beta_j' x_q}}, \text{ where } (\beta_2', \beta_3', ..., \beta_T')' . \tag{3-4}$$

3.1.4 Data Preparation

3.1.4.1 Data Sources

Several data sources are used in the current analysis. These sources include the following: a) vehicle classification counts conducted in the Dallas-Fort Worth area by the Texas Department of Transportation (TxDOT) Regional Planning Organization (RPO) and TxDOT's Transportation Planning and Programming Division; b) 1996 GIS-based road network file for the Dallas-Fort Worth area; c) zonal level land-use characteristics file of the Dallas-Fort Worth area; and d) 1996 GIS-based Dallas-Fort Worth zonal coverage file. The latter three data files were obtained from the North Central Texas Council of Governments (NCTCOG). Each of the four data sources is discussed briefly in the following paragraphs.

The TxDOT vehicle classification counts used in the analysis were conducted at several fixed stations in the Dallas-Fort Worth area during the periods from 1977 to 1987 and 1983 to 1993. The counts covered all the functional roadway classes and a mixture of land uses, with intent to obtain a sample that is representative of the VMT mix in the region. The counts were recorded using a manual count board from 6 a.m. to 10 p.m. on weekdays in the same month every year. The 16-hour counts were expanded to a 24-hour period to allow researchers to obtain the 24-hour vehicle classification counts. These 24-hour counts form the basis for computing the VMT mix on links. The counts distinguish among the following vehicle types: automobiles, pickups and vans (PUV), sports utility vehicles (SUV), combination trucks (two axles, three axles, four axles, and six axles), buses (two axles and three axles) and motorcycles (including all two-wheelers). The counts separate trucks by the number of axles, but we combined them for the current analysis because of the frequent occurrence of zero-counts in several axle categories. We also combined two-axle and three-axle buses into a single bus category for the same reason. The fraction of counts of each vehicle type represents the VMT mix at the individual link level and is the dependent variable in the current analysis.

The 1996 GIS road network file includes information on the characteristics of each link in the Dallas-Fort Worth metropolitan planning area. The metropolitan area (MA) includes about a 4,980 sq mi area with more than 45,000 unique roadway links to represent the roadway network. The MA covers the existing urbanized area and the contiguous area expected to be urbanized by the year 2020. It includes all of Collin, Dallas, Denton, Rockwall, and Tarrant counties and portions of Ellis, Johnson, Kaufman, and Parker counties. The link attributes available in the network file include length of the link, traffic direction, functional classification, number of lanes, free speed, capacity, and whether the link is divided.

The zonal level land-use characteristics file of the Dallas-Fort Worth area contains land-use data at the level of the traffic survey zone used by NCTCOG for its travel demand modeling purposes. There are about 6,000 traffic survey zones in the Dallas-Fort Worth MA. The land-use data for each zone include total land area and acreage in several individual land-use purposes (such as in manufacturing and warehousing; in retail, hotel, and motel; in institutional buildings like churches, government, museums, schools, and hospitals; and in airport runways and terminals).

The 1996 GIS-based Dallas-Fort Worth zonal coverage file provides the spatial configuration of the traffic survey zones in the Dallas-Fort Worth MA.

3.1.4.2 Data Assembly

The objective of the data-assembly steps was to append the appropriate link and zonal characteristics to each link observation. To accomplish this, we first spatially overlaid the 1996 GIS road network file and the 1996 GIS zonal coverage file. Next, each link at which vehicle counts were recorded was manually queried in the network database using the name of the street and the names of the cross streets at the end nodes. Once the link at which a count was made was spatially located in the GIS road network coverage, its identifying number in the network file was extracted. Also, the traffic analysis zone that spatially contains the link of interest was identified from the GIS zonal coverage. Using these link and zonal identifier fields, the relevant link and land-use characteristics were mapped to each vehicle count observation.

The raw TxDOT vehicle classification counts included 370 observations of link vehicle counts, of which only 244 observations could be geo-coded in the manner discussed above. These 244 link-count observations constituted the final sample for analysis. The vehicle type distribution in this final sample was almost the same as the vehicle type distribution in the raw data.[2]

3.1.5 Empirical Analysis

3.1.5.1 Sample Description

The descriptive statistics of the dependent variable (i.e., the fractional splits among the six vehicle types across observations) in the sample are provided in Table 3-1. As expected, on average, the automobile fractional split is highest, followed by the fraction of PUVs. The average ratio of SUVs is between 3 and 5 percent. However, at an individual link level, the SUV percentage is as high as 48.7 and the truck percentage is as high as 26.3. The fraction of buses and motorcycles in the vehicle stream is relatively low.

[2]The reader will note that even if the aggregate sample VMT mix does not reflect the actual aggregate VMT mix in a region, the estimated model parameters will still be consistent except for the category-specific constants. This is because of the multinomial logit structure of Equation 3-4. Consistent values of the category-specific constants can be obtained in a straightforward fashion if the aggregate vehicle type distribution in the local region is known (see Ben-Akiva and Lerman 1985)

Table 3-1. Fractional split of vehicle types

Vehicle Type	Mean Value	Std. Dev.	Minimum	Maximum
Autos	0.653	0.088	0.389	0.875
Pickups/Vans (PUV)	0.262	0.062	0.098	0.416
Sports Utility Vehs. (SUV)	0.035	0.034	0.000	0.487
Trucks	0.043	0.045	0.000	0.263
Buses	0.002	0.008	0.000	0.118
Motorcycles and Two-Wheelers (MC)	0.005	0.003	0.000	0.023

The percentage of observations for which the fractional mix of trucks, buses, and motorcycles is at or very close to the boundary value of 0 is rather high. In particular, the truck fraction is less than 0.01 for 33 percent of observations, the bus fraction is less than 0.01 for 99 percent of observations, and the motorcycle fraction is less than 0.01 for 95 percent of observations. Thus, using the log-odds analysis method (Equation 3-1) would be inappropriate in the current modeling context. The specification in Equation 3-2, which can accommodate boundary values of the dependent variable, is the appropriate approach.

Five sets of independent variables were included in the model to predict the VMT mix on links. These are a) link functional classification, b) link physical attributes, c) link free speed variables, d) degree of urbanization of the zone in which link lies, and e) zonal land-use characteristics. A number of variables within each of the five variable classes were considered in the model specification. The final set of variables and their method of inclusion in the VMT mix model was determined based on, first, a systematic process of eliminating variables found to be statistically insignificant in previous specifications, and, second, considerations of parsimony in representation. In the description below, we briefly highlight some of the characteristics of the variables in each of the five sets of variables that were retained in the final model specification.

The link functional classification identifies each link with one of five roadway classes: freeways, major arterials, minor arterials, collectors, and local or residential roads. Because the number of observations on local or residential roads was very small (only 4 out of the sample size of 244), we combined the collector and local/residential road classes into a single “collector/local” category. The sample split among the four resulting roadway classes is as follows: freeways (41.8 percent), major arterials (26.2 percent), minor arterials (13.5 percent), and collector/local links (18.5 percent).

Two link physical attributes turned out to be important determinants of the link VMT mix; these were the number of lanes and whether the link was divided. Of the links in the sample, a majority (56.6 percent) has two lanes; 10.2 percent of links have one lane, 24.2 percent have three lanes, and 9 percent have four lanes. A substantial percentage (82.4) of links are divided roads.

The link free speed varies between 9 mph and 68 mph, with a mean value of 45 mph. A direct specification with free speed as the independent variable was inferior to the specification that categorized links into one of four free speed groups: low speed (less than or equal to 30 mph), low-to-medium speed (31 to 40 mph), medium speed (41 to 55 mph), and high speed (greater than 55 mph).

The degree of urbanization of the zone in which the link lies is determined by classifying the zone as a central business district (CBD), an urban residential area, or a suburban or rural area (because the differentiation between suburban and rural areas did not affect the VMT mix, these two categories were combined). In the sample, about 5 percent of links are in a CBD area, about 40 percent in an urban area, and the remainder in a suburban or rural area.

The zonal land use variables include a) an airport presence variable indicating the presence or absence of airport runway or terminal facilities in the zone in which the link lies; b) an institution presence variable indicating the presence or absence of institutions such as churches, schools, and hospitals; c) zone acreage in retail and office space; and d) acreage in manufacturing and warehousing. In the sample, about 5 percent of links lie in a zone with airport-related infrastructure and about 49 percent of links are in zones with some land use for institutional facilities. The average zone acreage in retail and office space in the sample is 18.43 acres and the average zone acreage in manufacturing and warehousing is 31 acres.

3.1.5.2 Fractional Split Model Results

The final model specification results are provided in Table 3-2. The table provides estimates of the β parameter vector in Equation 3-4.

The link functional classification variables are introduced with the freeway class acting as the base roadway category. The results indicate an increase in the PUV and motorcycle fractions on major and minor arterials relative to freeways and to other vehicle types. The fraction of these two vehicle types, however, is highest on collector/local streets. The bus fraction is lower on minor arterials (compared to freeways and major arterials) and even lower on collector or residential streets.

Table 3-2. VMT fractional split model estimation results

Variable	Parameter Estimate	t-Statistic
Link functional classification		
Major arterials		
PUV	0.0934	2.090
MC	0.3595	4.130
Minor arterials		
PUV	0.1076	1.177
Bus	-1.0570	-2.193
MC	0.3138	2.506
Collector/local streets		
PUV	0.2416	3.427
Bus	-1.7264	-3.140
MC	0.6679	4.460
Link physical attributes		
Divided road		
Truck	1.1389	3.149
Bus	-0.6862	-2.328
MC	0.3427	2.625
Number of lanes		
Truck	-0.1738	-2.202
Bus	-0.5230	-2.045
Link free speed variables		
Low speed		
PUV	-0.2903	-3.838
SUV	-0.7688	-6.499
Truck	-1.7293	-7.013
Bus	1.0436	2.610
Low-to-medium speed		
PUV	-0.1469	-2.297
SUV	-0.3377	-3.259
Truck	-1.8454	-11.480
Bus	0.5063	1.744
Medium speed		
Truck	-0.4125	-3.847
MC	0.1481	1.829
Degree of urbanization		
Central Business District		
PUV	-0.2919	-3.205
Truck	-1.0350	-3.409
Bus	1.7342	4.106
Urban residential		
PUV	-0.0918	-1.932
SUV	-0.2322	-2.395
Truck	-0.5645	-4.209

Table 3-2. VMT fractional split model estimation results (cont.)

Zonal land-use variables		
Airport presence		
PUV	0.1823	2.211
Institution presence		
Auto	0.1207	2.904
Acreage in office/retail space		
PUV	-0.0019	-2.399
SUV	-0.0038	-1.796
Truck	-0.0165	-4.282
Bus	-0.0146	-1.937
MC	-0.0026	-2.169
Acreage in manufacturing/warehousing		
PUV	0.0009	3.427
SUV	0.0021	5.494
Truck	0.0067	9.632
Bus	0.0031	2.457

Notes: 1) PUV - Pickups and vans, SUV - Sports utility vehicle, MC - Motorcycles/two-wheelers

2) The estimated constant values for each vehicle type are as follows (the auto vehicle type is the base): -0.8147 (for PUV), -2.1601 (for SUV), -2.4148 (for trucks), -4.2927 (for buses), and -5.3752 (for MC).

The results of the effect of link physical attributes indicate an increase in truck fraction and a decrease in bus fraction on divided roads. Motorcycles are also more prevalent on divided highways than are other nontruck vehicle types. The impact of the number of lanes on vehicle mix suggests a decrease in the truck and bus fractions in the vehicle fleet on links with several lanes.

The link free speed variable coefficients show fewer PUVs and SUVs as a fraction of total vehicles on low-speed links relative to buses and passenger cars and relative to medium and high-speed links. The same, though more tempered, negative trend exists for PUVs and SUVs on low to medium-speed links. Generally speaking, PUVs and SUVs are more prevalent on higher-speed links than on lower-speed links. The same is true for the truck mix in the vehicle fleet, except that this effect is much stronger for trucks than for SUVs and PUVs. The results also indicate that the bus fraction is highest on low-speed facilities and is higher on low to medium-speed facilities than on high free-speed links. Finally, the fraction of motorcycles and other two-wheelers is higher on medium-speed links than on other links.

The coefficients on the variables characterizing the degree of urbanization show a lower fraction of trucks on links in CBD and urban residential zones relative to links in suburban/rural zones. Among the nontruck vehicle types, the PUV fraction is likely to be smaller than that of the other vehicle types on CBD links, and the SUV fraction is likely to be smaller than that of other vehicle types on urban links. In addition, the bus fraction is highest on CBD links compared to other link types.

The final set of variables is the set of land-use variables. The results reveal that the proportion of PUVs is high on links in zones with airport facilities. This finding is quite reasonable because PUVs are more convenient for transporting baggage and passengers to airports. The auto proportion is high on links in zones where institutions such as churches, schools, and hospitals are present. Similarly, the auto proportion is high on links in zones with large areas allocated to retail and office space. Finally, vehicle types other than automobiles and motorcycles are likely to capture a high proportion of the VMT mix in zones with large areas invested in manufacturing plants and warehouses.

3.1.6 Model Application

The model results in Table 3-2 can be applied in forecasting mode to determine the VMT mix in the six vehicle types: autos, PUVs, SUVs, trucks, buses, and motorcycles and two-wheelers. The implementation is particularly straightforward when one uses a GIS platform. This is the method the research team is using to determine the VMT mix on each link in the Dallas-Fort Worth MA as part of an ongoing air quality-related project funded by TxDOT.

The model-predicted VMT mix in the six vehicle types has to be converted into the eight-class EPA vehicle classification for input into the MOBILE emissions-factor model. We propose an approach that combines local vehicle registration data from the Dallas-Fort Worth area with information provided in the *Transportation Energy Data Handbook* (Davis 1997, p. 10) for this conversion, as discussed below.

The *Transportation Energy Data Handbook* estimates that 98.8 percent of passenger cars are gasoline-driven and 1.2 percent are diesel-driven. These splits are used to allocate the "autos" VMT mix between the LDGV (light-duty gasoline-powered vehicles) and LDDV (light-duty diesel-powered vehicles) EPA categories. Pickups and vans (PUVs, including minivans and passenger vans) and sports utility vehicles (SUVs) fall under the classification of light-duty trucks and are to be assigned among the LDGT1 (light-duty gasoline-powered trucks of gross vehicle weight less than 6,000 lb), LDGT2 (light-duty gasoline-powered trucks of gross vehicle weight between 6,000 and 8,500 lb), and LDDT (light-duty diesel-powered trucks of gross vehicle weight less than 8,500 lb) EPA vehicle types. From the *Transportation Energy Data Handbook,* we computed the gasoline-to-diesel split for light-duty trucks as 97.88 percent to 2.12 percent based on truck sale information up to 1995. This information is used to allocate 2.12 percent of the PUV VMT mix and the SUV VMT mix to the LDDT category. The remaining 97.88 percent of gasoline-powered PUVs and SUVs are allocated between the LDGT1 and LDGT2 categories based on 1996 local vehicle registration data obtained from TxDOT for the Dallas-Fort Worth region. Since the local vehicle registration data are differentiated by county, and the split of the LDGT1 and LDGT2 categories in the registration data are quite different across counties, we developed county-specific estimates of PUV and SUV splits in the LDGT1 and LDGT2 categories (our method, of course, assumes that the split of traffic in the LDGT1 and LDGT2 categories in each county is the same as the vehicle registration split in these categories in the county; the method does not consider intercounty travel, which may lead to a differential LDGT1/LDGT2 traffic split vis-à-vis the registration data split in each county) .

The 1996 local vehicle registration data for the Dallas-Fort Worth area provide the gasoline-diesel splits of combination trucks (vehicles with gross weight of over 8,500 lb) by county. This information is used directly to apportion the "combination trucks" VMT mix into the HDGV (heavy-duty gasoline vehicles) and HDDV (heavy-duty diesel vehicles) EPA categories.

A "bus" vehicle type classification is not included in the 1996 local vehicle registration data. Hence, we estimated the split of buses into the gasoline-powered and diesel-powered vehicles from the *Transportation Energy Data Handbook* (20.09 percent gasoline-powered and 79.91 percent diesel-powered) and used it to allocate the "bus" VMT mix between the HDGV and HDDV EPA vehicle categories.

Finally, the model-predicted "motorcycle" VMT mix is assigned completely to the MC EPA vehicle category.

Table 3-3 provides the final county-specific conversion factors between the six-vehicle type classification typology used in the VMT mix modeling of the current report and the eight-vehicle type EPA classification typology.

Table 3-3. TxDOT vehicle count vehicle type to MOBILE vehicle type conversion factors

Dallas County

TxDOT classification	EPA MOBILE vehicle type classification							
	LDGV	LDDV	LDGT1	LDGT2	LDDT	HDGV	HDDV	MC
Autos	98.8%	1.2%	-	-	-	-	-	-
PUV	-	-	95.16%	2.72%	2.12%	-	-	-
SUV	-	-	95.16%	2.72%	2.12%	-	-	-
Trucks	-	-	-	-	-	35.43%	64.57%	-
Buses	-	-	-	-	-	20.09%	79.91%	-
Motorcycles	-	-	-	-	-	-	-	100%

Tarrant County

TxDOT classification	EPA MOBILE vehicle type classification							
	LDGV	LDDV	LDGT1	LDGT2	LDDT	HDGV	HDDV	MC
Autos	98.8%	1.2%	-	-	-	-	-	-
PUV	-	-	96.07%	1.81%	2.12%	-	-	-
SUV	-	-	96.07%	1.81%	2.12%	-	-	-
Trucks	-	-	-	-	-	39.31%	60.69%	-
Buses	-	-	-	-	-	20.09%	79.91%	-
Motorcycles	-	-	-	-	-	-	-	100%

Collin County

TxDOT classification	EPA MOBILE vehicle type classification							
	LDGV	LDDV	LDGT1	LDGT2	LDDT	HDGV	HDDV	MC
Autos	98.8%	1.2%	-	-	-	-	-	-
PUV	-	-	96.15%	1.73%	2.12%	-	-	-
SUV	-	-	96.15%	1.73%	2.12%	-	-	-
Trucks	-	-	-	-	-	44.24%	55.76%	-
Buses	-	-	-	-	-	20.09%	79.91%	-
Motorcycles	-	-	-	-	-	-	-	100%

Table 3-3. TxDOT vehicle count vehicle type to MOBILE vehicle type conversion factors (cont.)

Denton County

TxDOT classification	**EPA MOBILE vehicle type classification**							
	LDGV	LDDV	LDGT1	LDGT2	LDDT	HDGV	HDDV	MC
Autos	98.8%	1.2%	-	-	-	-	-	-
PUV	-	-	96.36%	1.52%	2.12%	-	-	-
SUV	-	-	96.36%	1.52%	2.12%	-	-	-
Trucks	-	-	-	-	-	43.30%	56.70%	-
Buses	-	-	-	-	-	20.09%	79.91%	-
Motorcycles	-	-	-	-	-	-	-	100%

Rockwell County

TxDOT classification	**EPA MOBILE vehicle type classification**							
	LDGV	LDDV	LDGT1	LDGT2	LDDT	HDGV	HDDV	MC
Autos	98.8%	1.2%	-	-	-	-	-	-
PUV	-	-	95.96%	1.92%	2.12%	-	-	-
SUV	-	-	95.96%	1.92%	2.12%	-	-	-
Trucks	-	-	-	-	-	34.24%	65.76%	-
Buses	-	-	-	-	-	20.09%	79.91%	-
Motorcycles	-	-	-	-	-	-	-	100%

There are a few things to keep in mind about the conversion factors. First, these conversion factors can be updated continually as more local information becomes available. Second, the current EPA MOBILE emissions-factor model does not distinguish between PUVs and SUVs; both of these are classified as light-duty trucks. Thus, the distinction between PUVs and SUVs in our VMT mix model is rather academic at this point. However, new versions of the MOBILE models that distinguish between emissions of PUVs and SUVs are planned. The VMT mix model proposed here can be used to provide the disaggregate input needed by these forthcoming MOBILE models.

3.1.7 Model Evaluation

In this section, we evaluate the ability of the proposed model to replicate the actual VMT mix on links in the sample. We also compare the predicted emissions on each link (based on the proposed VMT mix model) with the actual emissions on that link (based on the observed link-specific VMT mix values). In addition, we compare the performance of the proposed model with that of a "default" model that uses only roadway functional classification as the controlling variable for VMT mix analysis (this is the state of the practice in the Dallas-Fort Worth area and in other metropolitan areas that use local VMT mix values). It will be noted that the "default" model is better than using the MOBILE5 default values because it is based on local conditions. But it is restrictive compared to the proposed model in this paper because it ignores factors other than roadway functional classification that may affect VMT fractions.

3.1.7.1 VMT Mix Performance Evaluation

For each sample link observation, we have data on the actual observed VMT mix in the six vehicle types: autos, PUVs, SUVs, trucks, buses, and motorcycles. We also have available for analysis the corresponding model-predicted VMT mix in these six vehicle types. The evaluation of the proximity of estimated and actual VMT mixes on links is based on three criteria: the mean absolute error (MAE), the mean percentage absolute error (MPAE), and a pseudo-R^2 value. The MAE is computed for each vehicle type as the average of the absolute difference between the model-predicted and actual VMT fractions for that vehicle type across link observations.

The MPAE is computed as the link-averaged absolute difference between the model-predicted and actual VMT fractions as a percentage of the actual VMT fraction. The pseudo-R^2 measure is an overall model-fit measure computed as shown below:

$$R^2 = \frac{\sum_q \sum_i \left(\hat{y}_{qi} - \bar{y}_i\right)^2}{\sum_q \sum_i \left(y_{qi} - \bar{y}_i\right)^2} , \tag{3-5}$$

In Equation 3-5 above, y_{qi} is the actual fraction of VMT accrued by vehicle type *i* on link *q*, $\hat{y}_{qi}$ is the model-predicted fraction, and $\bar{y}_i$ is the area-wide average VMT (from Table 3-1) for vehicle type *i*. The denominator in Equation 3-5 is the variation in the actual link VMT mix values around the mean areawide VMT mix value, summed across all vehicle types and links. The numerator represents the variation explained by the model. Thus, the pseudo-R^2 measure may be viewed as the fraction of total variation in VMT mix explained by the model. The measure varies between 0 and 1.

Table 3-4 provides the measures of fit (MAE, MPAE, and pseudo-R^2) for the proposed model and a "default" model that uses only roadway functional classification as the controlling variable. We do not compute an MPAE for buses and motorcycles in Table 3-4 because the actual VMT fractions for these vehicles are extremely small, leading to substantially high MPAE values by construction.

Table 3-4. VMT mix performance evaluation

Fit Statistic	Model	Vehicle Type					
		Auto	PUV	SUV	Trucks	Buses	Motorcycles
Mean Absolute Error (MAE)	Proposed	0.0482	0.0398	0.0133	0.0163	0.0016	0.0019
	Default	0.0720	0.0475	0.0150	0.0350	0.0018	0.0019
	% Higher error in default over proposed	49.38	19.35	12.78	114.72	12.50	0.00
Mean Percentage Absolute Error (MPAE)	Proposed	7.60	17.36	38.45	44.43	-	-
	Default	11.35	20.70	47.69	67.96	-	-
	% Higher error in default over proposed	49.34	19.24	24.03	52.96	-	-
Pseudo-R^2	Proposed	0.43					
	Default	0.03					

The MAE and MPAE results in Table 3-4 clearly indicate the superiority of the proposed model over the default model. The MPAE for SUVs and trucks are high even for the proposed model, but this is an artifact of very low VMT fractions of these two vehicle types on many links. The more significant observation is that there is a large improvement in the fit of the proposed model relative to the default model in these two vehicle classes, especially for trucks. More generally, the percentage higher error in the default model compared to the proposed model is quite substantial across all vehicle types.

The pseudo-R^2 measure of the proposed and default models in Table 3-4 is another indicator of the superior performance of the proposed model. The results indicate that the default model is able to explain only 3 percent of the variation of link VMT fractions, while the proposed model is able to explain 44 percent of this variation. This result implies that roadway classification alone does not contribute much to explaining VMT mix, and that there are several other very important link and land-use attributes that should be considered in VMT mix analysis. This implication is, of course, also quite apparent from the results in Table 3-2.

3.1.7.2. Emissions Performance Evaluation

To evaluate the benefit of the proposed VMT mix model for emissions prediction, we applied the conversion factors developed in Section 3.1.6 to convert the actual and predicted count-based VMT mix fractions into VMT fractions defined by the EPA typology. Since the total volumes and the lengths of each link in the sample were known, we computed the actual and predicted link VMT accrued by each of the EPA vehicle classes. Finally, we applied areawide emission factors that varied by pollutant type and EPA vehicle class (obtained from the MOBILE5 model by the North Central Texas Council of Government Staff for the Dallas-Fort Worth area) to calculate the “actual” and predicted emissions for each pollutant type on each link. The evaluation of the ability of the proposed model and the “default” model to replicate “actual” link emissions was based on the same three criteria used for VMT mix analysis, namely, mean absolute error (MAE), mean percentage absolute error (MPAE), and pseudo-R^2.

Table 3-5 provides the evaluation results. The MAE results in the table indicate the average link-level error in the emissions predictions (in grams), while the MPAE provides the average link-level percentage error in the emissions predictions. The results again indicate that the proposed model fits the data much better than does the default model. The MAE in the default model is about 24 to 30 percent higher than that obtained from the proposed model, while the MPAE in the default model is between 50 to 120 percent higher than that obtained from the proposed model, depending on the pollutant type (the MPAE in Table 3-5 for emissions is of an order lower than that in Table 3-4 for VMT mix because the magnitude of link emissions is very high compared to the VMT mix fractions; thus, a 1 percent error in emissions implies a much larger absolute error compared to a 1 percent error in VMT mix fractions).

Table 3-5. Emissions performance evaluation

Fit Statistic	Model	Pollutant Type		
		Carbon Monoxide (CO)	Volatile Organic Compounds (VOCs)	Oxides of Nitrogen (NOx)
Mean Absolute Error (MAE) in grams	Proposed	9992	684	6098
	Default	12754	853	8030
	% Higher error in default over proposed	27.64	24.76	31.70
Mean Percentage Absolute Error (MPAE)	Proposed	1.84	1.30	8.71
	Default	2.93	2.00	19.06
	% Higher error in default over proposed	59.24	53.40	118.82
Pseudo-R^2	Proposed	0.515		
	Default	0.004		

The pseudo-R^2 measure from the two models again emphasizes the superior performance of the proposed model. In summary, the improvement in VMT mix predictions by the proposed model does indeed translate to improved emissions estimation.

3.1.8 Conclusions

VMT mix, or the distribution of vehicles by weight and fuel type, is an important traffic-related parameter used in determining the composite mobile-source emissions on links of a network. The emissions factors (in grams per mile of vehicle travel for each pollutant) vary quite widely among different vehicle classes, and therefore the emissions analysis is very sensitive to the VMT mix. Consequently, it is important to develop methods that provide accurate VMT mix values at the individual link level.

Current approaches to VMT mix determination apply aggregate-level values across all links in the road network based on national-level traffic count statistics, or they apply roadway class-specific values based on local vehicle classification traffic counts. However, it has been documented in the literature that there is substantial variation in VMT mix across different regions and across links of the same roadway class within a region.

This paper proposes a fractional split model that relates VMT mix on links to several variables more informative and explanatory than merely roadway class. The fractional split model is a valuable formulation for VMT mix analysis because (1) it accommodates boundary values of the fractional VMT in a vehicle class; (2) it is easy to estimate using commonly available econometric software; and (3) it is easy to apply in forecasting mode to predict the VMT mix on each link of a network. A quasi-likelihood approach that provides consistent and asymptotically robust inference for the parameters in the fractional split model is used in estimation.

The empirical analysis in the paper applies the fraction split model structure to estimate a VMT mix model for the Dallas-Fort Worth MA in Texas. Several data sources are used to assemble the data needed in the estimation. This assembly requires a reasonable, though not very substantial, amount of effort. Once the data are assembled, estimation of the VMT mix model proposed here is straightforward, as is the application of the model to predict link VMT mix values. Thus, though the current paper uses the Dallas-Fort Worth region as the study area, similar models can be estimated easily in other areas. This is particularly the case today because many metropolitan areas now have network and land-use files in GIS format, from which the information required for the proposed model estimation can be immediately extracted.

The empirical results for the Dallas-Fort Worth area show important differences in VMT mix based on link functional classification, link physical attributes, link speed characteristics, the degree of urbanization of the zone that contains the link, and land-use variables of the zone in which the link lies. Model evaluation efforts indicate that the proposed model provides much better predictions of VMT mix and emissions estimation than does the default model in use by Metropolitan Planning Organizations. The proposed model is currently being embedded within a GIS platform to predict the VMT mix on all links of the Dallas-Fort Worth MA.

There are two limitations in the current empirical analysis. First, variations in VMT mix across different times of day are not captured in the model. Second, seasonal variations in VMT mix are not incorporated in our model. The vehicle-classification counts used in the current report provided only 24-hour counts and were conducted during the same month each year. Thus, they are inadequate for capturing temporal and seasonal variations. To accommodate these variations, more extensive vehicle-classification counts made at different times of day and different seasons of the year are needed. Once such data become available, the fractional split model structure can be applied to capture these additional effects. It is likely that the measures of fit of the proposed model in Tables 3-4 and 3-5 will improve even more after accommodating such temporal and seasonal variations in VMT mix.

3.2 MODELING TRIP DURATION FOR MOBILE SOURCE EMISSIONS FORECASTING

The emissions factor models take several traffic-related data as inputs, one of which is the distribution of the duration of vehicle trips in the region. The vehicle trip duration distribution is important for several reasons. First, the trip duration distribution provides information for developing trip duration activity parameters used by the MOBILE emissions factor model to estimate running loss emissions. Running loss emissions are evaporative emissions that have escaped from a vehicle while the engine is operating (from spots where the vehicle's evaporative/purge system has become inoperative). Due to greater heating of the engine fuel and evaporative system on longer trips, running loss emissions continually increase as a function of trip duration until the emissions reach a plateau at a trip duration of about 50 to 60 minutes (see Glover and Brzezinski 1998). Second, operating mode fractions, which are needed by MOBILE5 to estimate emissions rates, can be estimated from the trip duration distribution. Third, the trip duration distribution can be used to predict the vehicle miles of travel (VMT) accumulated on local roads in the region.

The modeling of trip durations in a metropolitan area is of considerable value for the reasons identified above. Trip duration is likely to depend on various factors such as the trip purpose, the time-of-day of the trip start, and other land-use and socio-demographic characteristics of the zone of trip start. In the current report, we formulate and implement a methodology for modeling trip durations using vehicle trip data from household travel surveys and supplementary zonal demographic/land-use data. The implementation is demonstrated in the context of mobile source emissions analysis for the Dallas-Fort Worth area in Texas.

3.2.1 Literature Review and Motivation for Study

Trip duration distribution is important for estimating running loss emissions, operating mode fractions and VMT on local roads. Correspondingly, we review the state-of-the-art/practice under these three headings: running loss emissions, operating mode fractions and VMT on local roads.

3.2.1.1 Running loss emissions

The methodology for estimating running loss emissions differs between MOBILE5 and MOBILE6. In MOBILE5, running loss emissions are modeled as a direct function of the input temperature, fuel volatility, and average speed. The procedure for calculating the running loss emissions entails partitioning the vehicle trip duration into six time duration bins (i.e., less than 10 minutes, 11 to 20 minutes, 21 to 30 minutes, 31 to 40 minutes, 41 to 50 minutes, and 51 minutes and longer) and obtaining the proportion of VMT accumulated by trips that fall into each time duration bin (these proportions are referred to as the trip duration activity parameters). Within MOBILE5, the running loss emissions value of an average vehicle trip is calculated as the sum of the product of the emission factors associated with each time duration bin (embedded within MOBILE5) and the corresponding trip duration activity parameter. The product of these average running loss emissions with the number of trips per day represents the running loss emission level. The user has the ability to accept default daily running loss emissions values available within MOBILE5 (developed using default trip-time distributions representing national average conditions), or develop region-specific estimates by specifying a local set of trip duration activity parameters. As a general recommendation, the MOBILE5 manual suggests using area-specific trip duration activity parameters to more accurately estimate running loss emissions.

MOBILE6 advances the state-of-the-art/practice by providing activity parameters for each of 14 time periods in a day and by distinguishing between weekdays and weekends. The default MOBILE6 hourly activity estimates are based on an EPA survey of 168 vehicles, and are invariant across geographic regions or across trip purpose categories. Thus, as in MOBILE5, EPA recommends the use of locally estimated trip duration activity parameters whenever possible.

In summary, using trip duration activity parameters developed from local data for estimating running loss emissions constitutes an important improvement over using the default values embedded in the MOBILE emissions factor model. In the current report, we present a methodology to develop zone-specific trip duration activity parameters that vary by time-of-day and trip purpose, using a trip duration model estimated from local data.

3.2.1.2 Operating Mode Fractions

Operating mode fractions are an important input to MOBILE5 in estimating mobile source emissions. There are two dimensions associated with operating mode fractions; one is the *start mode* of vehicle trips (cold versus hot), and the second is the *running mode* of vehicle trips (transient versus stabilized). Trip duration modeling, the focus of this report, affects the latter dimension of operating mode, i.e., the running mode of trips. To the extent that running mode fractions can be more accurately estimated using a trip duration model estimated from local data, such a model can contribute toward improved mobile source emissions forecasting.

EPA defines the *transient* mode of operation as all vehicle operations before 505 seconds after the start of a trip and the *stabilized* mode as all operations after 505 seconds of a trip. EPA recommends the following default values for running mode fractions: transient (47.9%) and stabilized (52.1%). The practice in most MPOs in the country is to accept these default running mode fractions. However, these default values were developed over twenty years ago and recent research suggests that it may no longer adequately represent overall vehicle emission control performance under current driving conditions. In addition, the default fractions do not vary by trip purpose, time-of-day, or regional land-use and socio-demographic characteristics.

Few studies have attempted to develop locally estimated running mode fractions of trips. Brodtmen and Fuce (1984) used field data obtained by direct on-road measurement of engine conditions to develop running mode fractions in New Jersey. Ellis et al. (1978) analyzed origin-destination data from travel surveys in Alabama to develop aggregate measures of running mode fractions. Frank et al. (2000) developed transient and stabilized mode fractions based on vehicle trip times, using the Puget Sound Panel Survey. This was one of the earliest studies that employed household-level travel survey information and land-use data for running mode fraction estimation. A study by Chatterjee et al. (1996) and Venigalla et al. (1999) used a network-based approach for modeling running modes. In these studies, the elapsed time of vehicles from trip origins was traced during traffic assignment of zone-to-zone trips on highway networks. The proportion of transient and stabilized modes on links were obtained by counting the number of trips assigned on each link that are of duration less than or greater than 505 seconds since their start. Allen and Davies (1993) have similarly used the ASSIGN module of MINUTP, a commercially available planning model, to determine trips operating in transient mode for the southern New Jersey area.

A limitation of the studies reviewed above is that they compute of a single set of running mode fractions for an entire state (or for aggregate regions within a state), and for various times-of-day and trip purposes. In this report, we estimate a trip duration model using local data from a metropolitan region and present a methodology to use this estimated model to develop running mode fractions that vary by zone within the region, time-of-day and trip purpose. In addition, our methodology allows for the estimation of running mode fractions for travel on local roads.

3.2.1.3 VMT on Local Roads

Local roads are usually not included in the travel demand model networks used by most MPOs, and hence the travel speeds and volumes required to calculate the VMT on local links are unavailable from travel demand models. Many MPOs simply calculate the VMT on local roads as a percentage (typically about 10%) of the VMT on all other roads, and use it in developing their emissions inventories. This method is rather ad hoc in nature, and can result in VMT estimates quite different from the actual values. A few MPOs model the VMT on local roads by attributing the local-road VMT separately to interzonal and intrazonal trips (see Chatterjee et al. 1997, p. 101, for a discussion). The VMT attributed to interzonal trips is modeled as the product of interzonal trips assigned (during traffic assignment) on centroidal link connectors and the coded length of the connectors. The VMT on local roads attributed to *intrazonal* travel is estimated as the product of the total intrazonal trips for each zone (obtained from the origin-destination trip-interchange matrices at the end of trip distribution), and an average intrazonal trip length parameter. This trip length parameter is typically calculated as a function of the total area of the zone. While this method is a substantial improvement over using a percentage of VMT on non-local roads, it is still limited by the restrictive nature of variation of the intrazonal trip length parameter. In particular, the intrazonal trip length (and, therefore, local VMT) does not vary by trip purpose, time-of-day and zonal spatial attributes (other than zonal area). Our study develops the intrazonal trip length as a function of time-of-day, purpose and zonal attributes. We accomplish this by estimating a trip duration model, and then multiplying the predicted intrazonal trip duration with an estimate of average speed on local links (it is more straightforward to develop a direct model of intrazonal trip length, but most household surveys collect data only on trip duration and not trip length).

In the next section, we present the model framework for the estimation and application of the trip duration model.

3.2.2 Model Framework

The modeling approach in the report uses vehicle trip data from household travel surveys and zonal demographic/land-use data from supplementary data sources. The approach involves developing the distribution of the duration of trips using a log-linear regression model. The use of a log-linear form for trip duration guarantees the non-negativity of trip time in application of the models.

The application step of the model predicts the distributions of the duration of trips for each traffic analysis zone in a metropolitan region, and for each combination of time-of-day and trip purpose. An important characteristic of the proposed method is the ease with which the estimated models from vehicle trip data can be immediately applied to obtain zonal-level trip-time distributions.

In the next section, we present the details of model estimation. In the subsequent section, we discuss the applications of the estimated model.

3.2.2.1 Model Estimation

Let q be the index for vehicle trip, t be the index for time-of-day, and i be the index for activity purpose prior to the trip. Define ω_{qti} to be a dummy variable taking the value 1 if vehicle trip q occurs in time-period t with trip purpose i, and 0 otherwise; define δ_{qz} as another dummy variable taking the value 1 if vehicle trip q originates in zone z, and 0 otherwise. Define I_q to be a variable that takes the value 1 if vehicle trip q is intrazonal, and 0 otherwise. Let x_z be a vector of zonal attributes.

We assume the trip duration to be log-normally distributed in the population of trips, and develop a linear regression model for the duration as a function of trip purpose, time-of-day and land-use and socio-demographic characteristics of the zone of trip origin.

Let d_q be the duration of vehicle trip *q*. Then, we write the log-linear regression equation for the trip duration as:

$$\ln(d_q) = \eta + \sum_{t,i} \alpha_{ti}\omega_{qti} + \lambda\left(\sum_{z} \delta_{qz} x_z\right) + I_q\left(\chi + \sum_{t,i} \zeta_{ti}\omega_{qti} + \rho\left(\sum_{z} \xi_{qz} x_z\right)\right) + \varepsilon_q, \varepsilon_q \sim N[0, \sigma^2] \quad (3\text{-}6)$$

In this equation, η is the generic constant to be estimated, α_{ti} (*t=1, 2,…T; i=1,2,…I*) are scalars to be estimated and λ is a vector of parameters also to be estimated. χ, ζ_{ti}, ρ and ξ_{qz} are similar to η, α_{ti}, λ and δ_{qz} respectively, but are introduced specific to intrazonal trips (note that I_q takes the value 1 if vehicle trip *q* is an intrazonal trip, and 0 otherwise). ε_q is a normally distributed random error term introduced to complete the econometric specification.

In Equation 3-6 above, we have not allowed interactions between zonal attributes and time-of-day/trip purpose combinations; however, this is purely for notational convenience and for ease in presentation of the model application step. Such interactions can be included within the model structure without any additional conceptual or estimation complexity. Similarly, the notation structure implies full interactions of time and trip purpose, though more restrictive structures such as single dimensional effects without interaction can be imposed by appropriately constraining the α_{ti} and ζ_{ti} scalars across the different time/trip purpose combinations.

The reader will note that the inclusion of the intrazonal dummy variable, and interactions of this variable with exogenous variables, allows us to accommodate separate trip duration distributions for intrazonal vehicle trips and interzonal vehicle trips.

The model from Equation 3-6 can be estimated using any commercially available software with a linear regression module. Data assembly issues for estimating the model are discussed in Section 3.2.3.

3.2.2.2 Model Application

This section discusses the application of the estimated model in the previous section. The subsequent three sections present the methodology to obtain 1) trip duration activity parameters for estimating running loss emissions, 2) running mode fractions of trips for use in MOBILE5, and 3) estimates of VMT for travel on local roads.

3.2.2.2.1 Trip duration activity parameters for running loss emissions

The trip duration distribution for any zone in the study area by time-period and trip purpose can be predicted in a straightforward manner after estimation of Equation 3-6. The (log) trip duration distribution of interzonal vehicle trips in time *t* for trip purpose *i* from zone *z* may be written as:

$$\ln\left(d_{tiz}^{a}\right) \sim N\left[\eta + \alpha_{ti} + \lambda x_{z}, \sigma^{2}\right] = N\left[\Delta_{tiz}^{a}, \sigma^{2}\right] \tag{3-7}$$

The mean Δ_{tiz}^{a} and variance σ^2 of this distribution can be estimated from the parameter estimates obtained in the estimation stage. The corresponding distribution of intrazonal vehicle trips in time *t* for trip purpose *i* in zone *z* may be written as:

$$\ln\left(d_{tiz}^{l}\right) \sim N\left[\eta + \alpha_{ti} + \lambda x_{z} + \chi + \zeta_{ti} + \rho x_{z}, \sigma^{2}\right] = N\left[\Delta_{tiz}^{l}, \sigma^{2}\right] \tag{3-8}$$

The objective in our effort is to obtain the fraction of VMT accrued by trips in each of six trip duration-bins (as needed by MOBILE) for each zone, and for each trip purpose and time-of-day combination. Let *k* be an index for time-bin (*k=1,2,..,6*), and let *k* be bounded by the continuous trip duration value of m_{k-1} to the left and by m_k to the right. Let V^k be the average speed of trips in time-bin *k* and let ϑ_z be the fraction of trips originating in zone *z* which are intrazonal.[3] Then, the fraction of VMT accrued by interzonal trips in time-bin *k* for in time *t* for trip purpose *i* originating in zone *z* ($FVMT_{tiz}^{ka}$) can be obtained as:

$$FVMT_{tiz}^{ka} = \frac{L_{tiz}^{ka} * \Omega_{tiz}^{ka} * V^{k}}{VMT_{tiz}^{a}} \tag{3-9}$$

where,

$$L_{tiz}^{ka} = \Phi\left(\frac{\ln\left(m^{k}\right) - \Delta_{tiz}^{a}}{\sigma}\right) - \Phi\left(\frac{\ln\left(m^{k-1}\right) - \Delta_{tiz}^{a}}{\sigma}\right) \tag{3-10}$$

---

[3] V^k may be obtained from local metropolitan area data or using the following national default values obtained from the 1995 National Personal Transportation Study (NPTS) data: 18.96 mph (for trips of duration 0-10 mins), 20.80 mph (for trips of duration 11-20 mins), 26.40 mph (for trips of duration 21-30 mins), 29.14 mph (for trips of duration 31-40 mins), 33.60 mph (for trips of duration 41-50 mins) and 45.30 mph (for trips of duration greater than 51 mins). ϑ_z represents the fraction of intrazonal trips originating from zone *z* and can be obtained from the sample used for estimation. If the sample data does not support evaluation of ϑ_z for all zones, ϑ_z can be determined from the zone-to-zone origin-destination trip interchanges matrices obtained at the end of the trip distribution step in the travel demand modeling process.

$$\Omega_{tiz}^{ka} = \exp\left[\Delta_{tiz}^{a} + \sigma \frac{\phi\left(\frac{\ln(m_{k-1}) - \Delta_{tiz}^{a}}{\sigma}\right) - \phi\left(\frac{\ln(m_{k}) - \Delta_{tiz}^{a}}{\sigma}\right)}{\Phi\left(\frac{\ln(m_{k}) - \Delta_{tiz}^{a}}{\sigma}\right) - \Phi\left(\frac{\ln(m_{k-1}) - \Delta_{tiz}^{a}}{\sigma}\right)}\right] \tag{3-11}$$

$$VMT_{tiz}^{a} = \sum_{k} L_{tiz}^{ka} * \Omega_{tiz}^{ka} * V^{k} \tag{3-12}$$

In the above equation structure, L_{tiz}^{ka} represents the proportion of interzonal trips in time period *t* for trip purpose *i* originating in zone *z*, that fall in trip-duration bin *k*. Ω_{tiz}^{ka} represents the mean trip duration of interzonal trips in time period *t* for trip purpose *i* originating in zone *z*, that fall in trip-duration bin *k*. The product of L_{tiz}^{ka} and Ω_{tiz}^{ka} with V^{k} represents the VMT accrued by interzonal trips in time period *t* for trip purpose *i* originating in zone *z*, that fall in trip-duration bin *k*. VMT_{tiz}^{a} represents the total VMT accrued by interzonal trips in time period *t* for trip purpose *i* originating in zone *z*, and is obtained by summing the VMT across all trip duration bins. Then, the proportion of VMT accrued by interzonal trips in time-bin *k* for time *t* for trip purpose *i* originating in zone *z*, $FVMT_{tiz}^{ka}$, is obtained as shown in Equation 3-9, by dividing the VMT accrued by trips in time-bin *k* by the total VMT.

The fraction of VMT accrued by intrazonal trips in time-bin *k* for time *t* for trip purpose *i* originating in zone *z* ($FVMT_{tiz}^{kl}$), can be obtained by substituting Δ_{tiz}^{l} instead of Δ_{tiz}^{a} in Equations 3-4 through 3-7.

Finally, the fraction of VMT accrued by all trips in each time-bin *k* for trip purpose *i* originating in zone *z* during time *t*, ($FVMT_{tiz}^{k}$), may be written as:

$$FVMT_{tiz}^{k} = \vartheta_{z} * FVMT_{tiz}^{kl} + (1 - \vartheta_{z}) * FVMT_{tiz}^{ka} \tag{3-13}$$

3.2.2.2.2 Running mode fractions for MOBILE5

This section presents the method to obtain the proportion of transient and stabilized trips required as an input to MOBILE5. We begin by discussing the approach for interzonal trips; the approach is identical for intrazonal trips, with appropriate replacements to reflect the mean and variance of intrazonal trips.

Let the assumed speed of vehicles be v. Let the mean of the distribution of trips of duration less than 8.42 minutes (505 seconds) occurring in time-period t with trip purpose i in zone z be μ_{tiz}^{1a} and let the corresponding mean of the distribution of trips of duration greater than 8.42 minutes be μ_{tiz}^{2a} (μ_{tiz}^{1a} and μ_{tiz}^{2a} represent the means of the right- and left-truncated normal distributions of trip duration respectively).

We obtain the analytical expression for μ_{tiz}^{1} (see Greene 1997) as:

$$\mu_{tiz}^{1a} = \Delta_{tiz}^{a} - \sigma \frac{\phi\left(\frac{\ln(8.42) - \Delta_{tiz}^{a}}{\sigma}\right)}{\Phi\left(\frac{\ln(8.42) - \Delta_{tiz}^{a}}{\sigma}\right)} \tag{3-14}$$

The VMT in transient mode accumulated by trips of duration less than (or equal to) 8.42 minutes, is given by $(\mu_{tiz}^{1a}) * (v) *$ [Number of trips of duration ≤ 8.42 min]. Trips of duration greater than 8.42 minutes are in the transient mode for the first 8.42 minutes of their operation. The VMT in transient mode accumulated by such trips is given by $\ln(8.42) * (v) *$ [Number of trips of duration > 8.42 min]. Therefore, the total VMT in *transient* mode is:

$$\left[(\mu_{tiz}^{1a}) * (v) * \Phi\left(\frac{\ln(8.42) - \Delta_{tiz}^{a}}{\sigma}\right) + \ln(8.42) * (v) * \left[1 - \Phi\left(\frac{\ln(8.42) - \Delta_{tiz}^{a}}{\sigma}\right)\right]\right] * \text{[Total number of trips]} \tag{3-15}$$

The mean duration of trips of duration greater than 8.42 minutes, μ_{tiz}^{2a}, is given by:

$$\mu_{tiz}^{2a} = \Delta_{tiz}^{a} + \sigma \frac{\phi\left(\frac{\ln(8.42) - \Delta_{tiz}^{a}}{\sigma}\right)}{1 - \Phi\left(\frac{\ln(8.42) - \Delta_{tiz}^{a}}{\sigma}\right)} \tag{3-16}$$

The VMT in stabilized mode in time t for trip purpose i originating in zone z can be obtained as $\left[\mu_{tiz}^{2a} - \ln(8.42)\right] * (v) *$ [Number of trips of duration > 8.42 min]. Therefore, the expression for the VMT accumulated in the *stabilized* mode is:

$$\left[\mu_{tiz}^{2a} - \ln(8.42)\right] * (v) * \left[1 - \Phi\left(\frac{\ln(8.42) - \Delta_{tiz}^{a}}{\sigma}\right)\right] * \text{[Total number of trips]} \tag{3-17}$$

The fraction of VMT in transient and stabilized modes can be obtained from Equations 3-6 and 3-8 for any zone *z*, and for any combination of time period *t* and trip purpose *i*, after substituting the estimated values of Δ^a_{tiz} and σ from the estimation stage. Thus, a distinguishing characteristic of the proposed method is the straightforward manner in which model parameters estimated from vehicle trip data can be applied to obtain zonal-level estimates of vehicle running mode fractions.

The reader will also note that the running mode fractions for intrazonal trips may be readily obtained using Equations 9 through 12 after replacing Δ^a_{tiz} with Δ^l_{tiz}, and using an average speed for v corresponding to local roads (we assume v = 20 mph for local roads).

3.2.2.2.3 VMT on local roads

As noted in the model estimation section, the intrazonal nature of a trip is captured through the interaction effects of I_q with exogenous determinants of trip duration. The logarithm of the trip duration of intrazonal trips in time *t* with trip purpose *i* in zone *z* is normally distributed, as shown in Equation 3-8. It follows from this that the trip duration distribution of intrazonal vehicle trips in time *t* with trip purpose *i* in zone *z* is log-normally distributed with a mean θ^l_{tiz} and variance λ^l_{tiz} given by the following expressions (see Johnson and Kotz 1970):

$$\theta^l_{tiz} = \exp\left(\Delta^l_{tiz} + \sigma^2/2\right) \tag{3-18}$$

$$\lambda^l = \exp\left(2 * \Delta^l_{tiz} + \sigma^2\right)\left[\exp\left(\sigma^2\right) - 1\right] \tag{3-19}$$

The mean trip length of intrazonal trips is the product of θ^l_{tiz} and the average speed on local roads (which we assume to be 20 mph). The total VMT on local roads due to intrazonal travel can next be estimated as the product of the mean intrazonal trip length and the total intrazonal vehicle trips (obtained from the trip-distribution step in the travel demand modeling process). Our methodology accommodates the variation in intra-zonal VMT on local roads with time-of-day, trip purpose and zonal socio-demographic and land-use characteristics through the variation of the average intrazonal trip duration with these characteristics.

3.2.3 Data Preparation

3.2.3.1 Data Sources

The data used in the empirical analysis were drawn from two sources: the 1996 Activity Survey conducted in the Dallas-Fort Worth (DFW) area and the zonal land use and demographics characteristics file for the DFW area. These data sources were obtained from the North Central Texas Council of Governments (NCTCOG).

3.2.3.2 Sample Formation

Several data assembly steps were involved in developing the sample. First, we converted the raw composite (travel and non-travel) activity file into a corresponding person-trip file. Second, we identified person-trips that were pursued using a motorized vehicle owned by the household. Third, we translated the person-trip file into a corresponding vehicle trip file, which provided the sequence of trips made by each vehicle in the household. In this process, we extracted and retained information on the time-of-day of each vehicle trip start, the traffic analysis process (TAP) zone of trip start location and trip-end location and the purpose of activity being pursued at the origin and destination of the trip. Fourth, we aggregated the TSZ-level (traffic survey zone, or TSZ, level; there are about 5,000 TSZs in the D-FW planning area) land-use and demographic characteristics to the TAP-level, and appended this information to each vehicle trip start based on the TAP in which the trip start occurs. Finally, we conducted several screening and consistency checks on the resulting data set from the previous steps (a flow chart of this screening process is available from the authors). As part of this screening process, we eliminated observations that had missing data on departure times, activity purposes, and/or on the TAP location of the vehicle trip start.

The final sample used for analysis includes 19,455 vehicle trip observations. Of these, 2,940 trips (15.1%) are intrazonal.

3.2.4 Empirical Analysis

3.2.4.1 Sample Description

The dependent variable of interest in our analysis is the time duration of trips. The trip duration for interzonal trips varies from a minimum of 1 minute to a maximum of 660 minutes (11 hours). The mean trip duration is about 21 minutes with a standard-deviation of about 24 minutes. The trip duration for intrazonal trips varies from a minimum of 1 minute to a maximum of 210 minutes (3.5 hours). The mean trip duration for such trips is about 11 minutes with a standard-deviation of about 18 minutes.

Three types of variables were considered to explain trip duration. These are: a) trip purpose variables indicating the purpose of the trip, b) time-of-day variables identifying the time of trip start, and c) zonal and trip attributes. Interactions among these three sets of variables were also considered. In the description below, we briefly highlight some of the characteristics of these sets of variables.

Trip purpose was characterized by two dimensions: whether or not the trip was produced at home (home-based versus non-home based trips) and the purpose at the attraction-end of the trip (*i.e.*, whether the attraction-end activity is work, school, social/recreational, shopping, personal business or other). Of the 19,455 trips, 14,294 (73.5%) are home-based. The distribution of intrazonal and interzonal trips by trip purpose is presented in Table 3-6.

Table 3-6. Distribution of trips by trip purpose

Trip Purpose	**Percentage distribution for**	
	Intrazonal	Interzonal
Home	73.4%	73.5%
Non-Home	26.6%	26.5%
Work	12.0%	22.3%
School	3.2%	2.4%
Social/Recreational	10.2%	11.3%
Shopping	6.9%	6.3%
Personal Business	40.0%	41.5%
Other	27.8%	16.2%

The trips are rather evenly spread across all attraction-end activity purposes for both intrazonal and interzonal trips. The percentage of work trips is higher for interzonal trips than for intrazonal trips, while the percentage of other trips is higher for intrazonal trips than for interzonal trips.

The time-of-day of variables were associated with one of the following six time-periods: morning (midnight-6:30 a.m.), a.m. peak (6:30 a.m.-9:00 a.m.), a.m. off-peak (9:00 a.m.-noon), p.m. off-peak (noon-4:00 p.m.), p.m. peak (4:00 p.m.-6:30 p.m.), and evening (6:30 p.m.-

midnight). The time-periods for the a.m. and p.m. peaks were based on the peak periods definitions employed by the transportation department of the NCTCOG in the D-FW area. The times for the off-peak periods were determined by splitting the remaining blocks of time at noon and midnight. The distribution of intrazonal and interzonal trips by time-of-day is presented in Table 3-7. In general, the distributions by time-of-day are rather similar across intrazonal and interzonal trips.

Table 3-7. Distribution of trips by time-of-day

Time-of-day of trip start	Percentage distribution for	
	Intrazonal	Interzonal
Morning	1.5%	4.0%
a.m. peak	22.4%	21.0%
a.m. off-peak	13.5%	12.8%
p.m. off-peak	28.4%	23.5%
p.m. peak	19.0%	22.6%
Evening	15.1%	16.1%

Several zonal (TAP-level) land-use and demographic characteristics were considered in our analysis. Of these, the following zonal attributes were significant determinants of trip duration: total zonal area, zonal household density, acreage in retail facilities, acreage in office space, number of people in service employment, acreage in institutional facilities (like hospitals, churches *etc*.), acreage in manufacturing and warehousing facilities, zonal median income and presence of airports or airport-related infrastructure in the zone. The trip-related attribute included in the model was an indicator variable for whether or not the trip was intrazonal.

The final model specification of trip duration was obtained by systematically eliminating statistically insignificant variables and combining those found to have similar and comparable effects in terms of magnitude and significance. The empirical results for the estimated model are discussed in the following section.

3.2.4.2 Results of Trip Duration Model

The empirical results for the log-linear regression model are presented in Table 3-8. The table provides the estimated values of $\eta, \alpha_{ti}, \lambda, \chi$ and ζ_{ti} (*t=1, 2,…T; i=1,2,…I*) in Equation 3.6.

Table 3-8. Empirical results for trip duration model

Variable	Coefficient	t-statistic
Constant	1.087	76.81
Trip purpose		
"Non-Home" purpose is base		
Home	0.093	16.16
"Work" purpose is base		
School	0.018	1.56
Social/Recreational	0.054	5.11
Shopping	-0.130	-16.07
Other	-0.093	-13.07
Time-of-day variables ("evening" period is base)		
morning - a.m. peak/p.m. peak	0.193	15.57
a.m. off-peak/p.m. off-peak	0.076	6.75
Time-of-day and trip purpose interaction effects		
morning - a.m. peak/p.m. peak x Non-Work	-0.067	-11.23
a.m. off-peak/p.m. off-peak x Social/Recreational	-0.113	-8.94
Zonal and trip-related attributes		
Zonal size-related variables		
Zonal Area x 10^{-6}	5.401	2.56
Zonal Acreage in Office Space x 10^{-3}	1.030	3.85
Number of People in Service Employment x 10^{-5}	1.102	9.75
Zonal Acreage in Manufacturing Facilities x 10^{-4}	3.001	5.19
Zonal Acreage in Retail Facilities x 10^{-4}	-9.847	-4.82
Zonal Acreage in Institutional Facilities x 10^{-4}	-4.434	-2.54
Zonal non-size-related variables		
Zonal Household Density x 10^{-4}	-7.008	-4.42
Median Income of Zone x 10^{-6}	-1.124	-6.55
Presence of an airport or airport-related infrastructure	2.327	2.56
Trip Related Variables		
Intrazonal trip	-0.337	-33.95
Intrazonal p.m. peak trip	-0.082	-5.08
Intrazonal Shopping, Social/Recreational trip	0.069	6.12
Number of observations	19,455	
Regression sums of squares	516.24	
Residual sums of squares	2,085.69	
R^2	0.198	
Adjusted R^2	0.198	

The trip purpose variables were included with non-home based trips as the base category (for home-based versus non-home based trips) and with work as the base attraction-end activity. The results indicate that home-based trips tend to be significantly longer than non-home based trips. Social/recreational trips are longer than work trips while shopping and other trips tend to be significantly shorter than work trips. These results are consistent with overall observed trends in household travel behavior (see Hu and Young 1999).

The time-of-day variables are introduced with the evening period being the base. The morning and a.m.-peak periods are combined into a single period because of very few trips in these periods (see Table 3-6). The time-of-day variables are statistically significant and intuitive in the direction of their effect on trip duration. In general, peak-period trips are longer in duration, followed by mid-day trips. The interaction effects of time-of-day and trip purpose suggest that non-work trips are of shorter duration during the peak periods relative to work trips, and social/recreational trips are of shorter duration than trips of other purposes during the off-peak periods.

Several zonal and other trip attributes have a statistically significant effect on trip duration. We classify these attributes into three categories: zonal size-related variables, zonal non-size related variables, and trip-related variables. The effects of these three sets of variables are discussed in the next three paragraphs.

Among the size-related variables, a larger total area of a zone, in general, increases the duration of trips originating in that zone. This is particularly the case if the zone has a high acreage in office space, perhaps reflecting the long return-home trips from work, and long non-work trips due to lower non-work activity opportunities from such zones. Similarly, trips originating in zones with a high number of people in service employment and with large acreage in manufacturing facilities also have longer durations. These may reflect congestion effects. On the other hand, acreage in retail and institutional facilities have a negative effect on trips duration, possibly due to greater accessibility to shopping and service-related activities in zones with higher retail and institutional acreage.

The zonal non-size related variables indicate smaller trip durations in zones with high household density and with high household income. However, trips originating in zones with an airport have a longer duration. This latter effect may reflect increased congestion effects on roadways in zones with airport-related infrastructure.

Finally, intrazonal trips are significantly shorter in duration than interzonal trips, especially during the p.m. peak, though the magnitude of this effect is less for shopping and social/recreational trip purposes.

3.2.5 Integration with Travel Demand Models

This section discusses issues related to integrating the trip duration model presented in this report with existing travel demand models. Existing travel demand models may be based on an activity approach or on a trip approach. Activity-based travel demand models focus on the activities that people pursue, as a function of the locations and attributes of potential destinations, the state of the transportation network, and the personal and household characteristics of individuals (see Ettema and Timmermans 1997). If such an approach is adopted in travel analysis, the activity stops made by individuals are explicitly modeled as a function of origin and destination activity categories, time-of-day, and zone of origin. Thus, information on trip purpose, time of trip start and attributes of the zone of trip origin are readily available for all trips. Integration of the trip duration model developed in this report within this framework is rather straightforward.

If a trip-based travel demand modeling framework is used, the trip duration model in the current report can be directly applied if the MPO develops zone-to-zone origin-destination interchanges for the disaggregate trip purpose and time-of-day categories identified in this report. However, most MPOs use more aggregate classes of trip purpose and time periods (typically home-based work, home-based other and non-home based trip purposes, and peak versus off-peak time periods). In this situation, the trip duration model can be used after post-processing the aggregate origin-destination trip interchanges matrix to reflect the disaggregate classifications employed here. Factors obtained from travel surveys can be applied to achieve this post-classification. In Tables 3-9, 3-10 and 3-11, we present such factors developed for the DFW region.

Table 3-9. Cross-classification of home-based work trips by time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Home-based work	9.02%	34.97%	6.33%	13.56%	26.57%	9.54%

Table 3-10. Cross-classification of home-based other trips by trip purpose and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Home-based school	0.66%	1.15%	2.42%	3.69%	4.57%	9.67%
Home-based social/recreational	0.12%	0.71%	3.07%	5.15%	4.93%	4.73%
Home-based shopping	0.29%	1.89%	4.41%	4.84%	3.78%	2.53%
Home-based personal business	0.64%	10.22%	2.20%	6.73%	6.09%	5.69%
Home-based other	0.06%	3.35%	0.94%	2.90%	1.66%	0.90%

Table 3-11. Cross-classification of non-home based trips by trip purpose and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Non-home based work	0.02%	1.10%	3.53%	4.05%	1.05%	0.27%
Non-home based school	0.00%	0.12%	0.21%	0.48%	0.41%	0.04%
Non-home based social/recreational	0.16%	0.76%	4.36%	10.13%	2.34%	2.56%
Non-home based shopping	0.06%	0.41%	2.07%	4.84%	3.60%	2.19%
Non-home based personal business	0.14%	1.72%	6.67%	10.11%	5.27%	2.13%
Non-home based other	0.41%	8.35%	3.08%	7.63%	6.68%	3.04%

3.2.6 Conclusions

The modeling of trip durations in a metropolitan area is important for the following reasons: First, trip duration activity parameters used by the MOBILE emissions factor model to estimate running loss emissions can be developed from the trip duration distribution. Second, the trip duration distribution provides information for estimating operating mode fractions, which are needed by MOBILE5 to estimate emissions rates. Third, the trip duration distribution can be used to predict the vehicle miles of travel (VMT) accumulated on local roads in the region.

Trip duration is likely to depend on various factors such as trip purpose, time-of-day of the trip start, and other land-use and socio-demographic characteristics of the zone of trip origin. In the current report, we formulate and implement a methodology for modeling trip durations as a function of these characteristics, using vehicle trip data from household travel surveys and supplementary zonal demographic/land-use data. The approach involves developing the distribution of the duration of trips using a log-linear regression model. The modeling framework is implemented in the context of mobile source emissions analysis for the Dallas-Fort Worth area of Texas.

The proposed model contributes significantly toward improved mobile source emissions forecasting by systematically developing area-specific estimates of running loss emissions, running mode fractions, and VMT on local roads. A distinguishing characteristic of the methodology is the straightforward manner in which model parameters estimated from vehicle trip data can be applied to obtain zonal-level trip duration distributions. The model can be integrated easily within various travel demand-air quality modeling frameworks.

3.3 MODELING SOAK-TIME DISTRIBUTION OF TRIPS FOR MOBILE SOURCE EMISSIONS FORECASTING: TECHNIQUES AND APPLICATIONS

As already indicated in the VMT and trip duration sections, the emissions factor models (MOBILE or EMFAC7F) require several traffic-related inputs. One of these inputs is the distribution of the soak-time of vehicle trip starts. For its MOBILE6 model, EPA has defined the soak-time as the duration of time in which the vehicle's engine is not operating and which precedes a successful vehicle start (a successful vehicle start is defined as a vehicle start that does not result in a stall). If the soak-time is less than 12 hours, the corresponding engine start is designated a "hot-start." If the soak-time is more than 12 hours, then the engine start is defined as a "cold-start".

MOBILE6 uses the soak-time for a vehicle trip as one of several factors to estimate the emissions associated with the engine start for the trip. The soak-time for a trip is likely to depend on various factors such as the activity purpose preceding the trip start (i.e., the origin activity purpose), the time-of-day of the trip start, and other land-use and socio-demographic characteristics of the zone of trip start. Comprehensive household travel surveys, conducted periodically by Metropolitan Planning Organizations (MPOs), can be used to extract information on soak-times and can form the basis for the development of models of soak-time duration as a function of the various factors identified above.

In the current report, we formulate and implement a methodology for modeling soak-time durations. The methodology involves estimation of models using vehicle trip data from household travel surveys and supplementary zonal demographic/land-use data. The effectiveness of the methodology lies in its easy application at the traffic zonal level within a metropolitan region to obtain zone-specific soak-time distributions by time-of-day and origin activity purpose. Data from a household travel survey conducted in the Dallas-Fort Worth area of Texas is used in the empirical analysis of the report.

3.3.1 State of the Art/Present Practice

3.3.1.1 MOBILE5 versus MOBILE6

MOBILE5 uses the concept of operating mode fractions in determining mobile source emissions. Specifically, the use of MOBILE5 requires the classification of vehicle miles of travel into three operating modes: cold-transient, hot-transient, and hot-stabilized. EPA defines the *transient* mode of operation as all operations before 505 seconds after the start of a trip and the *stabilized* mode as all operations after 505 seconds after the start of a trip. Transient trips are further classified as cold-transient or hot-transient depending on whether the start mode was a cold-start or a hot-start, respectively. For the MOBILE5 model, EPA has defined cold-starts to be starts that occur at least four hours after the end of the preceding trip for non-catalyst-equipped vehicles, and at least one hour after the end of the preceding trip for catalyst-equipped vehicles. Hot-starts are those that occur less than four hours after the end of the preceding trip for non-catalyst vehicles, and less than one hour after the end of the preceding trip for catalyst equipped vehicles. MOBILE5 recommends using the following percentages as default values for operating mode fractions: cold-transient (20.6%), hot-transient (27.3%), and stabilized (52.1%). The analyst using MOBILE5 also has the option of providing region-specific operating mode fractions. It is important to note here that operating mode fractions represent a combination of start modes (cold versus hot) and running mode (transient versus stabilized).

In MOBILE6, the “start” emissions are estimated separately from the “running” emissions (emissions emitted while the vehicle is being driven). The start emissions are calculated as emission “increments” resulting from vehicle start-ups while the running emissions estimates are based only on hot-stabilized engine operations. The procedure for calculating the start emissions entails partitioning the hot vehicle starts (ranging from 1 minute to 12 hours) into 70 time-bins and assigning an emissions effect to each of the time-bins. From the distribution of soak-times, the proportion of soaks that fall into each time-bin is obtained. Within MOBILE6, the emission value of an average vehicle start is calculated as the sum of the product of the start emission effects associated with each time-bin and the corresponding soak-length activity proportion. The product of this average vehicle start emissions with the number of starts per day represents the start emission level. Estimates of hourly start-emission level values developed using default soak-time distributions representing national average conditions are available within MOBILE6. The user has the ability to accept these default emission level values or develop region-specific estimates by specifying a local distribution of soak-times.

In the next section, we review some of the extant methodologies for modeling start emissions and discuss the distinguishing characteristics of the current study.

3.3.1.2 Present Practice

The literature on modeling start emissions focuses on estimating start mode fractions (i.e., cold versus hot starts, as defined by MOBILE5) rather than developing the distribution of soak-times. This is because MOBILE5 requires start mode fractions (as part of developing the operating mode fractions) and not soak-time distributions. As indicated earlier, MOBILE6 advances the practice by emphasizing the underlying disaggregate soak-time distribution for emissions estimation. The authors are not aware of any prior study that models soak-time duration from regional travel data. But, to the extent that the start mode fractions used by MOBILE5 are aggregate representations of soak-time distribution, we will review earlier studies on start mode fractions in this section.

The practice in most MPOs in the country is to accept the default start (and operating) mode fraction values developed by EPA through its Federal Test Procedure (FTP). However, these default values were developed over twenty years ago and recent research (see EPA Report 420-R-93-007 Office of Mobile Sources, U.S. EPA [1993]) suggests that it may no longer adequately represent overall vehicle emission control performance under current driving conditions. In fact, as a general recommendation, EPA suggests the use of locally estimated values of traffic and other inputs whenever possible.

Few studies have attempted to develop locally estimated start mode fractions of trips. Brodtmen and Fuce 1984) used field data obtained by direct on-road measurement of engine conditions to develop start mode fractions in New Jersey. Ellis et al. (1978) analyzed origin-destination data from travel surveys in Alabama to develop aggregate measures of start modes. In another study, Garmen Associates obtained the percentage of cold-starts in North and South New Jersey using a combination of the MOBILE default values and a few local test runs. These methods, however, have the following limitations: a) they make ad hoc and strong assumptions regarding start modes by trip purpose (for example, the study by Garmen Associates assumes that all work trips are cold-starts); b) they involve substantial and expensive field data collection; and c) they compute a single set of values to be applied throughout a state or spatially aggregate regions within a state.

More recently, Venigalla et al. (1995) used data from the National Personal Transportation Survey (NPTS) to model cold and hot-start proportions as a function of trip purpose and time-of-day. The study indicated that trip purpose and time-of-day are the most important variables affecting the variation in start mode fractions. Venigalla et al.'s study is an important contribution to estimating start mode proportions, and highlights the value of using travel survey data to determine start mode fractions.

In this study, as in the report by Venigalla et al., we use household travel survey data to determine soak-time durations. However, there are important differences between our approach and that of Venigalla et al. First, our focus is on the disaggregate soak-time distributions, while that of Venigalla et al. was on the more aggregate start mode fractions. The approach we develop can be used to provide the soak-time distributions needed by MOBILE6 or can be aggregated to provide the start fractions needed by MOBILE5. Second, we examine soak-time durations as a function of zonal land-use and socio-demographic characteristics in addition to time-of-day and purpose. Our analysis, therefore, allows the development of distinct soak-time duration distributions for each zone in a metropolitan area. Venigalla et al.*'s* analysis is at a more geographically aggregate level, allowing variations across metropolitan areas in the country based on urban area size, but not allowing local variations within an urban area. Third, we explicitly distinguish soak-time distributions for first starts of the day and non-first starts of the day, since the distributions for these two types of starts are likely to be very different. Fourth, our purpose taxonomy is based on the activity pursued prior to the trip start rather than the trip purpose taxonomy used by Venigalla et al. To highlight this difference, consider a set of trips from home to shop and the corresponding reverse set of trips from shop to home. In the approach used by Venigalla et al., both trips would be classified as home-based other and would be assigned the same start mode fractions. In our approach, we allow the possibility that the soak-time for the trip starts from home to shop is higher than for the trip starts from shop to home, since the soak-duration is likely to be a function of the activity purpose being pursued during that soak-time (in our empirical analysis, we also examined the effect of destination activity purpose of the trip on soak-time duration, but did not find any significant impact of this variable). Finally, our analysis allows us to determine separate soak-time distributions for interzonal trips and intrazonal trips. Assuming that most intrazonal trips are on local streets, the latter distribution provides the zone-specific soak-time distributions for use with local road vehicle miles of travel (VMT) estimates.

3.3.2 Model Framework

The modeling approach in the report uses vehicle trip data from household travel surveys and zonal demographic/land-use data from supplementary data sources. Three steps are involved in model estimation. The first step models whether a vehicle trip start is the first of the day or not, using a discrete binary logit model. The second step analyzes the soak-time distribution for the set of first vehicle trip starts of the day, using a log-linear regression model. The third step models the soak-time distribution for the set of non-first trip starts of the day, also using a log-linear regression model. The use of a log-linear form for soak-time guarantees the non-negativity of soak-time in application of the models.

The application step of the model predicts the soak-time distributions of trip starts for each traffic analysis zone in a metropolitan region, and for each combination of time-of-day and activity purpose. An important characteristic of the proposed method is the ease with which the estimated models from vehicle trip data can be immediately applied to obtain zonal-level soak-time distributions.

In the next section, we present the details of model estimation. In the subsequent section, we discuss model application to obtain zone-specific soak-time distributions for each combination of time-of-day and activity purpose.

3.3.2.1 Model Estimation

Let q be the index for vehicle trip start, t be the index for time-of-day, and i be the index for activity purpose prior to the trip. Let G be an indicator variable taking the value 1 if a trip start is a first start and 0 otherwise. Define ω_{qti} to be a dummy variable taking the value 1 if trip start q occurs in time-period t and is preceded by activity purpose i, and 0 otherwise; define δ_{qz} as another dummy variable taking the value 1 if trip start q occurs in zone z, and 0 otherwise. Let x_z be a vector of zonal attributes.

We use a binary logit formulation to model whether or not a trip start is the first start of the day. The probability that trip start q is the first of the day is written as:

$$P_q(G=1) = \frac{1}{1+e^{-\left[\left(\sum_{t,i}\gamma_{ti}\omega_{qti}\right)+\beta'\left(\sum_{z}\delta_{qz}x_z\right)\right]}} \qquad (3\text{-}20)$$

where γ_{ti} $(t=1,2,\ldots T;\ i=1,2,\ldots I)$ are scalars, and β is a vector of parameters indicating the effect of zonal attributes on the probability of a first start.

Next, we assume the soak-time to be log-normally distributed in the population of trip starts, and develop separate linear regression models for first starts and non-first starts. Let a_q^F be the soak-time for trip start q if it is a first start, and let a_q^{NF} be the soak-time for trip start q if it is a non-first start. Then, we write the log-linear regression equations for first starts and non-first starts as follows:

$$\begin{aligned}
\ln\left(a_q^F\right) &= \eta^F + \sum_{t,i}\alpha_{ti}^F\omega_{qti} + \lambda^F\left(\sum_{z}\delta_{qz}x_z\right) + \varepsilon_q^F,\ \varepsilon_q^F \sim N\left[0,\left(\sigma^F\right)^2\right] \\
\ln\left(a_q^{NF}\right) &= \eta^{NF} + \sum_{t,i}\alpha_{ti}^{NF}\omega_{qti} + \lambda^{NF}\left(\sum_{z}\delta_{qz}x_z\right) + \varepsilon_q^{NF},\ \varepsilon_q^{NF} \sim N\left[0,\left(\sigma^{NF}\right)^2\right]
\end{aligned} \qquad (3\text{-}21)$$

where α_{ti} (*t=1, 2,…T; i=1,2,…I*) are scalars to be estimated, λ^F and λ^{NF} are vectors of parameters also to be estimated, and ε_q^F and ε_q^{NF} are normally distributed random error terms introduced to complete the econometric specification.

In the equation structures of 3-20 and 3-21 above, we have not allowed interactions between zonal attributes and time-of-day/activity purpose combinations; however, this is purely for notational convenience and for ease in presentation of the model application step. Such interactions can be included within the model structure without any additional conceptual or estimation complexity. Similarly, the notation structure implies full interactions of time and activity purpose, though more restrictive structures such as single dimensional effects without interaction can be imposed by appropriately constraining the γ_{ti}, α_{ti}^F and α_{ti}^{NF} scalars across the different time/activity purpose combinations. Finally, we also include an intrazonal dummy variable, and interactions of this variable with time-of-day/origin purpose, in our empirical specification. This allows us to accommodate separate soak-time distributions for intrazonal vehicle trips and interzonal vehicle trips. Again, for simplicity in presentation, we suppress these additional intrazonal variables in the presentation of model structure.

3.3.2.2 Model Application

This section discusses the application of the estimated models in the previous section to obtain the soak-time distribution for use in MOBILE6 and to obtain the start mode fraction of trips for use in MOBILE5.

3.3.2.2.1 Soak-time distribution for MOBILE6

Once the model parameters in Equations 3-20 and 3-21 are estimated from disaggregate vehicle trip start data, the soak-time distribution for any zone in the study area by time-period and activity purpose can be determined in a rather straightforward manner. To see this, first consider the binary logit model for first starts. The probability that any trip start *q* in time-period *t* preceded by an activity of purpose *i*, and occurring in zone *z*, is a first start can be written as:

$$P_{tiz}(G=1) = \frac{1}{1+e^{-\left[\gamma_{ti}+\beta' x_z\right]}} \tag{3-22}$$

Let M_{tiz} be the total number of trip starts in time-period *t* preceded by an activity of purpose *i*, and occurring in zone *z*. Then, the fraction of first trip starts in time *t* with purpose *i* in zone *z* can be written as:

$$\delta_{tiz} = \frac{M_{tiz} * P_{tiz}(G=1)}{M_{tiz}} = P_{tiz}(G=1) \tag{3-23}$$

Hence, once the γ_{ti} and β parameters are estimated, the fraction of first starts in time *t* with purpose *i* in zone *z* can be computed using the expression in Equation 3-22.

Next, the (log) soak-distribution of trip starts in time *t* with origin activity purpose *i* in zone *z* for first starts and non-first starts may be written as:

$$\begin{aligned} \ln\left(a_{tiz}^{F}\right) &\sim N\left[\eta^{F} + \alpha_{ti}^{F} + \lambda^{F} x_{z}, \left(\sigma^{F}\right)^{2}\right] = N\left[\Delta_{tiz}^{F}, \left(\sigma^{F}\right)^{2}\right] \\ \ln\left(a_{tiz}^{NF}\right) &\sim N\left[\eta^{NF} + \alpha_{ti}^{NF} + \lambda^{NF} x_{z}, \left(\sigma^{NF}\right)^{2}\right] = N\left[\Delta_{tiz}^{NF}, \left(\sigma^{NF}\right)^{2}\right] \end{aligned} \tag{3-24}$$

The means (Δ_{tiz}^{F} and Δ_{tiz}^{NF}) and variance ($\left(\sigma^{F}\right)^{2}$ and $\left(\sigma^{NF}\right)^{2}$) of these distributions can be estimated from the parameter estimates obtained in the estimation stage.

The objective in our effort is to obtain the fraction of soaks in each of 70 time-bins (as needed by MOBILE6) across both first- and non-first starts for each zone, and for each activity purpose and time-of-day combination. Let *k* be an index for time-bin (*k=1,2,…,70*), and let time-bin *k* be bounded by the continuous soak-time value of m_{k-1} to the left and by m_k to the right. Then, the fraction of soaks in time-bin *k* for first starts may be written as:

$$fraction\left(m^{k-1} < a_{tiz}^{F} < m^{k}\right) = \Phi\left[\frac{\log\left(m^{k}\right) - \Delta_{tiz}^{F}}{\sigma^{F}}\right] - \Phi\left[\frac{\log\left(m^{k-1}\right) - \Delta_{tiz}^{F}}{\sigma^{F}}\right] \tag{3-25}$$

The corresponding expression for non-first starts is:

$$fraction\left(m^{k-1} < a_{tiz}^{NF} < m^{k}\right) = \Phi\left[\frac{\log\left(m^{k}\right) - \Delta_{tiz}^{NF}}{\sigma^{NF}}\right] - \Phi\left[\frac{\log\left(m^{k-1}\right) - \Delta_{tiz}^{NF}}{\sigma^{NF}}\right] \tag{3-26}$$

Finally, the fraction of soaks in time-bin *k* across first and non-first starts for zone *z*, time-of-day *t* and activity purpose *i* may be computed as follows:

$$\begin{aligned} fraction\left(m^{k-1} < a_{tiz} < m^{k}\right) = {} & fraction\left(m^{k-1} < a_{tiz}^{F} < m^{k}\right) * \delta_{tiz} + \\ & fraction\left(m^{k-1} < a_{tiz}^{NF} < m^{k}\right) * \left(1 - \delta_{tiz}\right) \end{aligned} \tag{3-27}$$

3.3.2.2.2 Start mode fractions for MOBILE5

MOBILE5 requires the mode fractions in cold-starts and hot-starts to compute the operating mode fractions. As indicated in Section 3.3.1.1, hot-starts are defined as those with a soak-time of 1 hour or less for vehicles with catalytic converters and with a soak-time of less than 4 hours for vehicles without catalytic converters. If a start is not a hot-start, it is a cold-start. Thus, the fraction of hot-starts among the population of trip starts made by vehicles with catalytic converters for zone *z*, time-of-day *t* and original activity *i* is as follows:

$$\textit{fraction(hot-starts)} = \Phi\left[\frac{\log(60) - \Delta_{tiz}^{F}}{\sigma^{F}}\right] * \delta_{tiz} + \Phi\left[\frac{\log(60) - \Delta_{tiz}^{NF}}{\sigma^{NF}}\right] * (1 - \delta_{tiz}) \qquad (3\text{-}28)$$

The fraction of cold-starts among the population of trip starts made by vehicles with catalytic converters for zone *z*, time-of-day *t* and original activity *i* is trivially obtained as *[1-fraction(hot-starts)]*.

The expression for the fraction of hot-starts among the population of trip starts made by vehicles not equipped with catalytic converters is the same as Equation 3-28 with log(240) substituted for log(60).

3.3.3 Data Preparation

3.3.3.1 Data Sources

The data used in the empirical analysis are drawn from two sources: the 1996 Activity Survey conducted in the Dallas-Fort Worth (DFW) area and the zonal land use and demographics characteristics file for the DFW area. These data sources were obtained from the North Central Texas Council of Governments (NCTCOG), and are briefly discussed next.

The 1996 activity survey collected information on activities undertaken during a weekday by members of 4,839 households. For non-travel activities, information on the activity type, start and end times of participation, and location was collected (the location of each activity was geo-coded to a traffic analysis process, or TAP, zone; there are 919 TAP zones in the DFW planning area). For travel activities, information on the mode of travel used, costs incurred, and trip duration was collected. In addition, the survey elicited individual and household socio-demographic information.

The zonal level land use and demographics characteristics file contained land use and demographic data at the level of the traffic survey zone (TSZ) within the D-FW metropolitan planning area. The land use information for each TSZ provides information on total land area and acreage in several individual land use purposes (for example, in manufacturing; in retail, hotel and motel; in institutional buildings such as churches, government, museums, schools, and hospitals; in multifamily households; and in airport runways/terminals). The demographic information for each TSZ includes population, number of households, population density, median income, average household size, etc.

3.3.3.2 Sample Formation

Several data assembly steps were involved in developing the sample. First, we converted the raw composite (travel and non-travel) activity file into a corresponding person-trip file. Second, we identified person-trips that were pursued using a motorized vehicle owned by the household. Third, we translated the person-trip file into a corresponding vehicle trip file, which provided the sequence of trips made by each vehicle in the household. In this process, we extracted and retained information on the time-of-day of each vehicle trip start, TAP zone of trip start location and trip end location, purpose of activity being pursued during soak-time, and soak-time prior to vehicle trip start. The first trip start in the day for each vehicle was also identified and flagged. To compute the soak-times for these first trip starts, we assume that the soak-time prior to the first trip start is invariant across days for each vehicle (this assumption is necessary because only a single day of diary data is available). The soak-time for the first trip starts can then be computed as the difference in time between the first trip start of the day and the last trip end of the diary day. Fourth, we aggregated the TSZ-level land-use and demographic characteristics to the TAP-level, and appended this information to each vehicle trip start based on the TAP in which the trip start occurs. Finally, we conducted several screening and consistency checks on the resulting data set from the previous steps (a flow chart of this screening process is available from the authors). As part of this screening process, we eliminated observations that had missing data on departure times, activity purposes, and/or on the TAP location of the vehicle trip start.

The final sample used for analysis includes 18,231 vehicle trip start observations. Of these, 4,246 (23.3%) were first starts, and 13,985 (76.7%) were non-first starts.

3.3.4 Empirical Analysis

3.3.4.1 Sample Description

The dependent variable of interest in model estimation is the soak-time duration distribution of trip starts. The soak-time duration for first starts varies from a minimum of 71 minutes to a maximum of 1,339 minutes (around 22 hours). The mean soak-time for first starts is 835 minutes (about 14 hours) with a standard deviation of about 204 minutes. The soak-time duration for non-first starts varies from a minimum of 1 minute to a maximum of 1,345 minutes (around 22 hours). The mean soak-duration for non-first starts is 161 minutes (about 2.5 hours) with a standard deviation of 196 minutes.

Three types of variables were considered to explain soak-time for first starts and non-first starts. These are: a) time-of-day variables identifying the time of trip start, b) activity purpose variables indicating type of activity pursued prior to the trip start, and c) zonal and trip attributes. Interactions among these three sets of variables were also considered. In the description below, we briefly highlight some of the characteristics of these sets of variables.

The time-of-day of trip start was associated with one of the following six time-periods: morning (midnight-6:30 a.m.), a.m. peak (6:30 a.m.-9:00 a.m.), a.m. off-peak (9:00 a.m.-noon), p.m. off-peak (noon-4:00 p.m.), p.m. peak (4:00 p.m.-6:30 p.m.), and evening (6:30 p.m.-midnight). The time-periods for the a.m. and p.m. peaks were based on the peak period definitions employed by the transportation department of the NCTCOG in the DFW area. The times for the off-peak periods were determined by splitting the remaining blocks of time at noon and midnight. The distribution of first starts and non-first starts by time-of-day is presented in Table 3-12. As can be observed, a substantial fraction of first starts occur in the a.m. peak period, reflecting the morning commute trip. Only a small fraction of first starts occur outside the a.m. periods. On the other hand, most of the non-first starts occur in the p.m. periods, reflecting a combination of return-home trips from work and trips to/from other non-work activities.

Table 3-12. Distribution of trip starts by time-of-day

Time-of-day of trip start	Percentage distribution for	
	first starts	non-first starts
morning	14.1%	0.5%
a.m. peak	64.4%	6.9%
a.m. off-peak	14.2%	12.6%
p.m. off-peak	5.7%	30.0%
p.m. peak	1.2%	28.9%
evening	0.4%	21.2%

The distribution of first starts and non-first starts by activity purpose prior to the trip start is provided in Table 3-13. About 96% of the first starts begin from home (analysis of the destination activity of these “first start” trips indicates that about 60% are destined to work or school, while the remainder are quite evenly distributed across other activity purposes). For non-first starts, the activity purpose prior to the trip start is much more evenly distributed, though close to half of all starts begin at work or home (about 47% of these “non-first start” trips are destined to home, while the remainder are rather evenly distributed across the other purpose categories).

Table 3-13. Distribution of trip starts by activity purpose preceding trip start

Activity purpose preceding trip start	Percentage distribution for	
	first starts	non-first starts
Home	96.4%	19.3%
Work	2.2%	29.9%
School	0.0%	3.3%
Social/Recreational	0.8%	15.4%
Shopping	0.1%	10.5%
Personal Business	0.3%	11.2%
Other	0.2%	10.4%

Several zonal (TAP-level) land-use and demographic characteristics were considered in our analysis. Of these, the following zonal attributes were significant determinants of the trip start type (first versus non-first) and/or soak-time duration: acreage in multifamily households, retail employment and service employment, number of households, and population density. The trip-related attribute included in the model was an indicator variable for whether or not the trip corresponding to the vehicle start was an intrazonal trip. Including this intrazonal trip indicator enables the distinction of soak-time duration for intrazonal (local) trip starts and interzonal trip starts. Of the 18,231 trips in the sample, 2,676 (14.7 %) are intrazonal.

3.3.4.2 Empirical Results

This section presents the empirical results for the estimated models. Section 3.3.4.2.1 discusses the estimation results for trip start type (first versus non-first starts). Section 3.3.4.2.2 presents the results for soak-time duration of first starts, and Section 3.3.4.2.3 presents the corresponding results for non-first starts.

3.3.4.2.1 Results for first starts vs. non-first starts

The binary logit model results for first starts vs. non-first starts are provided in Table 3-14. The base category used is non-first starts. Thus, a positive coefficient on a variable indicates that the variable increases the probability of a first start, while a negative coefficient implies that the variable decreases the probability of a first start. The constant in the model does not have any behavioral interpretation; it adjusts for the range of zonal attribute values in the sample and the sample shares of first starts.

Table 3-14. Empirical results for binary logit model for first starts

Variable	Coefficient	t-statistic
Constant	5.186	24.13
Time-of-day variables (Morning period is base)		
a.m. peak	-2.465	-11.736
a.m. off-peak	-4.251	-19.96
p.m. off-peak	-5.932	-27.55
p.m. peak	-7.069	-30.67
evening	-7.780	-30.61
Activity purpose prior to trip start ("Home" purpose is base)		
Work	-3.677	-30.19
School	-4.807	-8.99
Social/Recreational	-4.734	-21.48
Shopping	-5.728	-13.87
Personal Business	-5.402	-20.35
Other	-6.874	-23.20
Zonal and trip attributes		
Population x 10^{-5}	-6.949	-3.68
Number of households x 10^{-4}	1.709	3.58
Intrazonal trip	-0.496	-5.60
Number of Observations	18,231	
Log-Likelihood Function	-3196.08	
Log-Likelihood for Constants only	-10977.08	

The time-of-day variables are introduced into the model with the morning period as the base time-period. The signs on the estimated coefficients for all other time-periods are negative and increasing in magnitude from the a.m. peak to the evening. This implies that trip starts that occur earlier in the day are more likely to be first starts than those made later in the day.

The activity purpose variables are introduced using the "home" purpose as the base category. The purpose dummy variables for all other activities are negative, indicating that, everything else being equal, trips from home are most likely to be first starts. A comparison of the magnitudes of coefficients across the activity purpose categories provides additional information regarding the likelihood of first starts among the group of non-home trip starts. Specifically, starts from work are more likely to be first starts than from other non-home purposes; starts after school and social-recreational activities are more likely to be first starts than those after shopping, personal business, and other (for example, drop-off/pick-up) activities.

Among the zonal and trip attributes, the population of the zone of trip start, the number of households, and an indicator of whether the trip was an intrazonal one have significant impacts on the likelihood of the trip start being the first of the day. The effect of zonal population may reflect more opportunities for participation in activities (such as shopping, social-recreational, etc.) in highly populated areas, which would result in more trips made per vehicle. As the number of trips per vehicle increases, the fraction of first trip starts has to decline, which would explain the negative effect of zonal population on the likelihood of first starts. On the other hand, for a given zonal population size, a higher number of households would imply more spreading of trips across households. This can lead to a decrease in number of trips per vehicle, resulting in a greater likelihood of first starts. Finally, intrazonal trips are short-distance trips, and stops pursued in such trips may be more likely to be linked with (and pursued after) stops to more distant locations (such as a shopping stop near home on the way back from work). Consequently, starts for intrazonal trips are less likely to be first starts in the day.

The log-likelihood values at convergence and with only the constant are provided toward the bottom of Table 3-14. The hypothesis that the time-of-day, activity purpose, and zonal/trip variables have no impact on the probability of first starts is strongly rejected by a log-likelihood ratio test. The pseudo-R^2 value for the model is 0.71.

3.3.4.2.2 Results for soak-time duration model for first starts

The results for the soak-time duration model for first starts are given in Table 3-15. The dependent variable in the model is the logarithm of soak-duration.

Table 3-15. Empirical results for soak-time duration model for first starts

Variable	Coefficient	t-statistic
Constant	2.827	689.71
Time-of-day variables (Morning period is base)		
a.m. peak	0.069	15.71
a.m. off-peak	0.163	29.11
p.m. off-peak	0.225	30.63
p.m. peak/evening	0.270	22.06
Activity purpose and associated interactions ("Home" purpose is base)		
Work	-0.091	-8.80
Morning x "non-home/non-work" activity purpose	-0.481	-10.03
Zonal and trip attributes and associated interactions		
Acreage in multi-family households x 10^{-4}	-4.078	-3.64
Intrazonal trip	-0.288	-9.84
Intrazonal x Home origin	0.307	10.38
Number of observations	4,246	
Regression sums of squares	20.24	
Residual sums of squares	38.67	
R^2	0.344	
Adjusted R^2	0.342	

The time-of-day variables are introduced with the morning period being the base. The p.m. peak and evening periods are combined into a single period because of very few first starts in these periods (see Table 3-12). The sign and magnitudes of the time-of-day variables in Table 3-15 indicate that the soak-time for first starts occurring later in the day is higher than for those occurring earlier in the day.

The activity purpose categories are collapsed into three categories for modeling soak-time duration for first starts: home, work, and non-home/non-work. This aggregation is necessary because very few first starts are preceded by a non-home or non-work activity. Dummy variables for work and non-home/non-work purposes are included with home being the base activity. The sign on the work purpose variable in Table 3-15 suggests that first starts from work have a smaller soak-duration than those from home. This is quite intuitive, since the first starts from work presumably represent the early morning return-home trips of individuals who came in to work late the previous night. We did not find any generic impact of the "non-home/non-work" purpose variable across time periods; however, the interaction effect of this variable with the morning period is statistically significant. The sign on this interaction suggests that the soak-time duration for first trip starts in the early morning period after non-work and non-home activities (*i.e.,* after social-recreation, shopping, etc.) is smaller than for trip starts in the early morning period from home or work.

Among the zonal and trip attributes, acreage in multifamily households is a significant determinant of soak-time duration. The sign on this variable suggests a smaller soak-duration for first starts occurring in zones with large acreage in multifamily households. This may be a consequence of return-home trips in the early morning after a social event in the neighborhood. The intrazonal dummy variable and its interaction with the "home" purpose indicate that the soak-duration of first starts associated with intrazonal trips is smaller than the soak-duration of first starts associated with interzonal trips for trips not originating at home. However, the reverse relationship holds for trips originating at home. The last few rows of the table provide summary fit statistics. The R^2 value for the model is 0.34.

3.3.4.2.3 Results for soak-time duration model for non-first starts

The results for the soak-time duration model for non-first starts are presented in Table 3-16. The dependent variable in the model is the logarithm of soak-duration.

The time-of-day variables are introduced with the morning and a.m. peak period being the base. The morning and a.m. peak periods are combined into a single period because of very few non-first starts in these periods (see Table 3-12). The results indicate that the soak-time for non-first starts occurring later in the day is higher than for those occurring earlier in the day.

Table 3-16. Empirical results for soak-time duration model for non-first starts

Variable	Coefficient	t-statistic
Constant	1.667	45.88
Time-of-day variables (morning-a.m. peak period is base)		
a.m. off-peak	0.086	1.82
p.m. off-peak	0.171	4.51
p.m. peak	0.258	6.79
evening	0.291	7.64
Activity purpose prior to trip start ("Home" purpose is base)		
Work/School	-0.220	-3.42
Social/Recreational	-0.514	-11.96
Shopping	-0.777	-17.81
Personal Business	-0.974	-22.54
Other	-1.214	-29.63
Time-of-day and activity purpose interaction effects		
a.m. off-peak		
Work/School	0.402	5.45
Social-Recreational/Shopping/Personal Business/Other	0.311	5.75
p.m.		
Work/School	0.713	10.95
Social-Recreational/Shopping/Personal Business/Other	0.283	6.59
Zonal and trip attributes		
Number of people in retail and service employment x 10^{-5}	1.004	7.26
Intrazonal trip	-0.121	-8.85
Intrazonal "home" trip	0.082	3.01
Number of observations	13,985	
Regression sums of squares	4,279.89	
Residual sums of squares	3,452.17	
R^2	0.55	
Adjusted R^2	0.55	

The activity purposes are introduced with home as the base activity. Interaction effects of activity purpose with time-of-day are also introduced. For these interaction effects, we use aggregate classifications for time-of-day and activity purpose, based on extensive empirical testing. Specifically, we classify non-home purposes into two broad categories: work/school and non-work/non-school (the home purpose is the base). We also use only two broad time periods: a.m. off-peak and p.m. (the morning/a.m. peak period is the base).

Referring to Table 3-16, the main effect for the work/school purposes, considered with the interaction effect for work/school purposes with time-of-day, indicates the following: a) for the morning/a.m. peak periods, non-first starts from work/school have a smaller soak-duration than for non-first starts from home; and b) for other time periods, non-first starts from work/school have a larger duration than non-first starts from home. These results are intuitive, since individuals are likely to spend longer durations at work/school than at home during the day. The main effects for all other non-work and non-school purposes indicate that non-first starts from those purposes are associated with a smaller soak-duration than non-first starts from home, school and work. However, this is less so as the day progresses, presumably because individuals have more time to pursue shopping, personal business, social-recreational and other activities in the latter part of the day.

Among the zonal and trip attributes, the number of people in retail and service employment is a significant determinant of soak-time duration. The sign on this variable suggests a larger soak-duration for non-first starts occurring in zones with a large number of people in retail and service employment. The number of people in service and retail employment here may be viewed as indicator variables for the "size" of shopping and service-related opportunities in a zone; the bigger the "size," larger is likely to be the activity duration of participation and therefore, the soak-duration. The intrazonal dummy variable and its interaction with the "home" purpose suggest that the soak-duration of non-first starts is larger for intrazonal trips originating at home, but lesser when originating at non-home activities.

The last few rows of the table provide summary fit statistics. The R^2 value for the model is 0.55.

3.3.5 Application Considerations

The models estimated above can be applied as discussed in Section 3.3.4.2.3 to obtain zone-specific soak-time distributions for each combination of time-of-day and activity purpose. In this section, we discuss how these soak-time distributions can be used in combination with travel demand models.

Travel demand modeling may be based on an activity-based approach or on a trip-based approach (see Bhat and Koppelman 1999 for a detailed discussion). In the activity-based modeling approach, the emphasis is on activities, and trips are considered as the derivative of the need to participate in activities at different locations. The activity-based approach treats time as an all-encompassing continuous entity within which individuals make activity/travel participation decisions (see Kurani and Lee-Gosselin 1996). Thus, if an activity-based approach is used in travel demand modeling, the stops (and, therefore, vehicle trip starts) by purpose type, time-of-day, and zone of origin are modeled explicitly. This information can be immediately used with the zone-specific soak-time distributions by time-of-day and activity purpose developed in this report.

The methodology developed here can also be used with a trip-based approach. In the trip-based approach, the trip-interchanges (zone-to-zone production-attraction matrices) within each trip purpose category are determined first and are subsequently converted into a zone-to-zone origin-destination matrix by time-of-day (fixed factors are typically used in this conversion, though departure time choice models can be used to better serve this purpose; see Steed and Bhat 2000). Depending on the trip purpose classification used, the zone-to-zone origin destination matrix by time-of-day may or may not provide all the information needed for integration with the soak-time model developed in this report. Specifically, if the home-based trips are classified into home-based work, home-based shopping, home-based school, home-based personal business, home-based social-recreational, and other home-based trips in the trip-based modeling, then information on trip starts by zone of origin, activity purpose prior to the trip start in the classification scheme used in the soak-time model, and time-of-day is available for home-based trips. In addition, if the non-home-based trips are also further sub-classified by origin activity categories in the six non-home activity typology used in the soak-time model, then information on trip starts by zone of origin, activity purpose, and time-of-day is available for all trips, and this can be used with the soak-time distribution model.

Most MPOs do not use the level of disaggregation in trip purposes discussed earlier. In fact, many continue to use only three trip purpose types: home-based work, home-based other, and non-home-based (there is, however, an increasing trend toward using more disaggregate trip purpose categories). Until MPOs use a more disaggregate trip-based classification scheme, or use an activity-based approach, an alternative is to post-classify trips into the origin activity purpose and time-of-day adopted in our soak-time model. This can be done by applying fixed factors obtained from travel surveys. In Tables 3-17, 3-18, and 3-19, we disaggregate home-based work, home-based other, and non-home-based trips by activity at origin end and by time-of-day for the DFW region. These fractions can be applied to zone-to-zone production-attraction matrices in each of the three broad trip purpose categories to obtain the number of trip starts by zone of origin, activity purpose prior to trip start, and time-of-day. The implementation is particularly straightforward using a Geographic Information System (GIS) platform. This is the method that the research team is using to determine zone-specific soak-time distributions in the DFW metropolitan planning area as part of an ongoing air quality-related project funded by TxDOT.

Table 3-17. Cross-classification of home-based work trips by origin-end activity and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Home	7.82%	33.13%	4.64%	5.09%	1.35%	1.22%
Work	0.94%	0.79%	1.67%	8.79%	26.08%	8.49%

Table 3-18. Cross-classification of home-based other trips by origin-end activity and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Home	1.36%	13.91%	7.66%	9.77%	8.69%	8.30%
School	0.01%	0.03%	0.38%	2.12%	0.84%	0.81%
Social/Recreational	0.30%	0.29%	1.05%	2.62%	2.09%	8.93%
Shopping	0.06%	0.24%	1.51%	3.29%	3.53%	3.28%
Personal Business	0.13%	0.18%	1.75%	2.72%	2.85%	1.73%
Other	0.10%	2.04%	0.45%	2.51%	2.92%	1.55%

Table 3-19. Cross-classification of non-home-based trips by origin-end activity and time-of-day

Trip purpose	Time-of-day of trip start					
	morning	a.m. peak	a.m. off-peak	p.m. off-peak	p.m. peak	evening
Work	0.02%	1.39%	8.63%	13.42%	10.43%	2.06%
School	0.00%	0.16%	0.59%	1.50%	0.22%	0.28%
Social/Recreational	0.24%	0.99%	2.91%	9.46%	1.80%	4.22%
Shopping	0.04%	0.46%	1.94%	4.36%	2.16%	1.58%
Personal Business	0.14%	1.68%	4.97%	6.51%	2.69%	1.07%
Other	0.34%	6.67%	1.31%	2.55%	2.16%	1.05%

3.3.6 Conclusions

The temporal distribution of the engine-off (soak) times of trips in a region is an important traffic-related input to mobile source emissions models. The soak-time associated with a trip could depend on various factors such as the activity purpose preceding the trip start, the time-of-day of the trip start, and possibly other land-use and socio-demographic characteristics of the zone of trip origin. More accurate region-specific estimates of the distribution of soak-time and the operating mode fractions can be obtained by modeling the soak-time as a function of these variables. In the current report, we estimate a model of soak-time durations as a function of these different attributes. The modeling approach in the report uses vehicle trip data from household travel surveys and zonal demographic/land-use data from supplementary data sources. Three steps are involved in model estimation. The first step models whether a vehicle trip start is the first of the day or not using a discrete binary logit model. The second step analyzes the soak-time distribution for the set of first vehicle trip starts of the day with a log-linear regression model. The third step models the soak-time distribution for the set of non-first trip starts of the day, also using a log-linear regression model. Our proposed model framework contributes toward improved mobile source emissions modeling by developing a systematic approach to analyzing soak-time durations. The effectiveness of our methodology lies in its easy application at the traffic zonal level within a metropolitan region to obtain zone-specific soak-time distributions by time-of-day and origin activity purpose. We apply the model framework to obtain zone-specific soak-time distributions for the DFW region. However, it should be applicable to any metropolitan region after appropriate re-estimations of the trip-level soak-time models.

Notwithstanding the very general nature of our model formulation, it is important to acknowledge two limitations of the application of the formulation in the current report. First, the model does not capture seasonal variations in soak-time distribution or its variation between weekdays and weekends. This is because the travel diary survey in the DFW area was limited to a single survey day, and hence information on trip-making behavior of individuals across different days/seasons is not available. If and when such data become available, these day-to-day and seasonal variations can be accommodated in a straightforward manner within the framework of this report. Second, since the survey focused on *household* travel, no commercial trips are included in the survey. Extending the empirical application in this report to model the soak-time durations of commercial trips is an important direction for future research. Such an extension will require collection of survey data on commercial vehicle activity.

CHAPTER FOUR - TRANSCAD IMPLEMENTATION

4.1 INTRODUCTION

Geographic Information System (GIS) technology has the capability to assemble, store, display, and analyze geographically referenced information and has been widely used in transportation planning. The models developed in this study are integrated in TransCAD, a transportation GIS package. TransCAD forms the platform for the travel demand forecasting and air quality procedure to evaluate the effectiveness of TCMs.

In the current project, seven models are implemented in TransCAD: an ordered probit model for trip generation, a disaggregated attraction-end choice model for trip distribution, multinomial logit models for mode choice and departure time choice, a VMT mix model, a trip time duration model, and a soak-time duration model. All the procedures involve the use of Caliper Script^{TM}, a macro programming language used in the Geographic Information System Develops Kit (GISDK) of TransCAD. The integration procedures are discussed in detail in the following sections.

4.2 TRIP GENERATION MODEL IMPLEMENTATION

4.2.1 Introduction

As indicated in Chapter Two, the ordered response probit model structure is adopted to estimate the number of trips made by households for each trip purpose. First, the models were developed for home-based work trips (HBW), home-based non-work trips (HBNW), and non-home-based trips (NHB). Models with various specifications were tested; the final models included only statistically significant demographic variables. Second, the models were developed for disaggregate trip purposes that include home-based (HB) grocery shopping, HB non-grocery shopping, HB social, HB recreational, HB personal business, and HB community trips. All the models were estimated using econometric software packages.

4.2.2 Input File Descriptions

Two types of inputs are required for forecasting trip productions from each TAP. One is the estimated parameters from the ordered response choice model, while the other set of inputs are the joint distributions of the demographic variables for each TAP. For the DFW metropolitan area, the joint distributions of household size and income quartile are available at the TAP level.

However, our models include independent demographic variables such as the number of workers in households and age structure of households. Therefore, multidimensional joint demographic distributions must be obtained for forecasting purposes. As described in the travel demand modeling section, we developed a procedure to generate the synthetic population and therefore obtain the joint distribution of household demographics.

The input data, including the model parameters and the joint household demographic distributions, should be in dBase format. In the model parameter file, the independent variables should be listed first, followed by the threshold parameters characterizing the ordered response trip production relationship. Furthermore, the names of independent variables in the parameter file should be the same as those in the demographic distribution file. In the model parameter file, the threshold parameters should be labeled as MU1, MU2,…, MU*n*. Figure 4-1 shows an example of input files. In Figure 4-1(a), a model parameter input file is shown. In this example the only independent variable in addition to a constant is the number of workers (WORKER). The constant of the estimated model should appear in the first column and be labeled as "CONST", followed by all the other independent variables. The threshold parameters appear at the end of the file in ascending order. In this example five threshold parameters are included, indicating that the maximum number of trips is six.

The household demographic distribution file is shown in 4-1(b). The first column of the file should be TAP ID number. The last column should be the number of households in various demographic groups. The rest of variables correspond to the independent variables in the parameter file. For example, the demographic distribution file shown in 4-1(b) indicates that there are 69 households in TAP 4 without any household members employed, 125 households with one household member employed, and so forth.

Dataview1 - HBW_1

CONST	WORKER	MU1	MU2	MU3	MU4	MU5
-0.74	1.18	0.43	1.40	1.68	2.42	2.70

Figure 4-1 (a). Model parameter inputs

Dataview2 - tap_hhs

TAP	WORKER	NUM_OF_HHS
4	0	69
4	1	125
4	2	119
4	3	17
4	4	3
4	5	0
4	6	0
5	0	93
5	1	168
5	2	160
5	3	23
5	4	4
5	5	1
5	6	0

Figure 4-1(b). Household demographic distributions

Figure 4-1. Input files

4.2.3 Program Interface and Output

The TransCAD does not have a built-in function to estimate an ordered response probit model. We built a new procedure to implement the ordered response probit model in TransCAD. The output of the ordered response probit model is the number of trip productions by purpose for households in each TAP. The total zonal productions can be obtained by aggregating trips by purpose across households in each TAP. We developed add-in procedures in GISDK for each of these steps, which are described below in the context of home-based work trip productions.

1. Open a map (tsz90.map) that contains zonal structure information for the study area.
2. Go to menu “tools” and choose “add-ins”. In the “add-ins” window (as shown in Figure 4-2), choose “Ordered-response Model Forecasting”, and click “OK”. The implementation procedure is now activated.

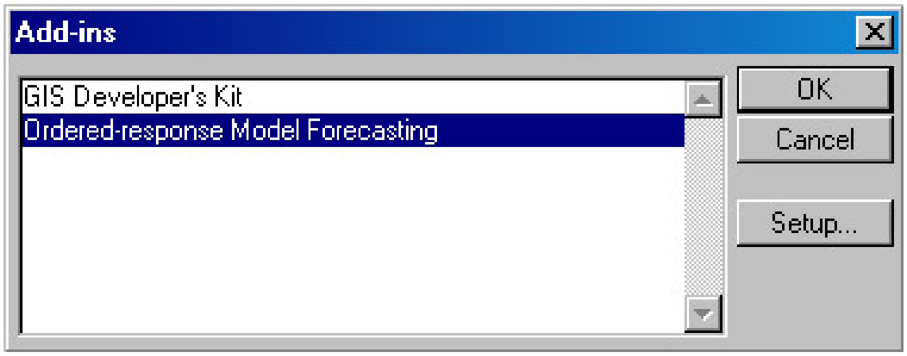


Figure 4-2. Add-in window

3. After the add-in procedure is activated, an open-file window (as shown in Figure 4-3) will pop out. First, select a household demographic distribution table (hbw_d.dbf) and click “open”; then, select a model coefficient table (hbw_1.dbf) and click “open”.

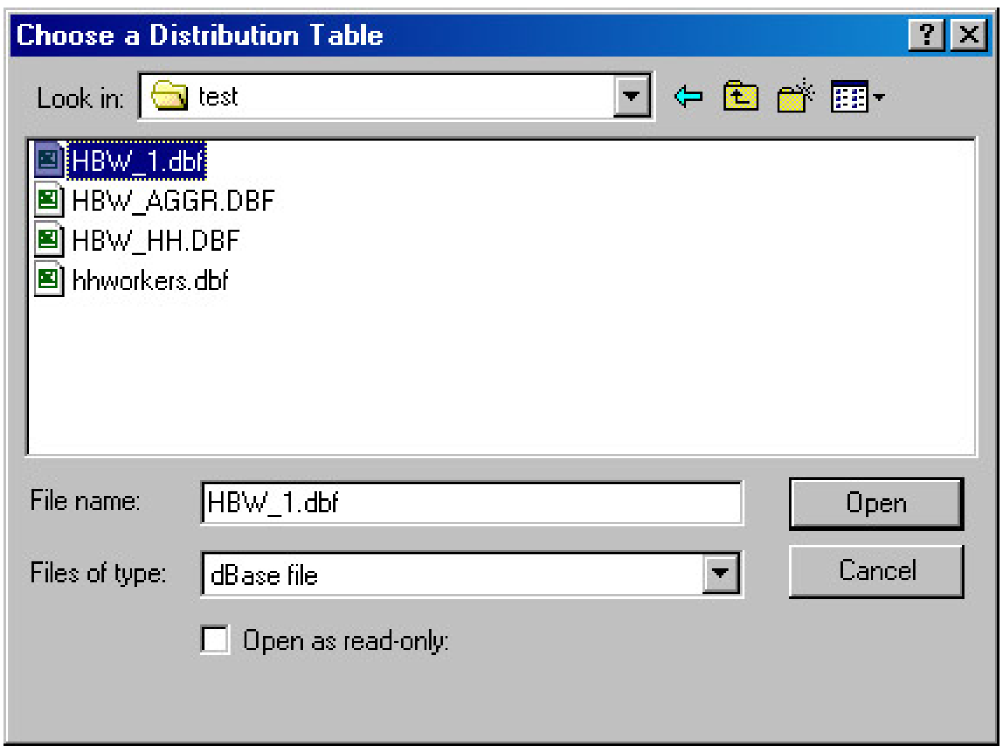


Figure 4-3. Open file window

4. An input dialogue box will then ask users to provide information on the number of independent variables in the model coefficient file, the maximum number of trips, and the trip purpose. The user can either choose inputs from a pull-down menu or type inputs

directly. In this example, the number of independent variables is 2; the maximum number of trips is 6; and the trip purpose is HBW (as shown in Figure 4-4).

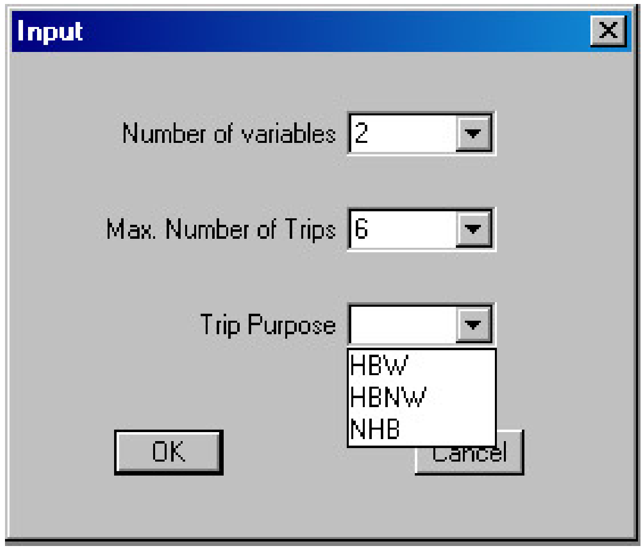


Figure 4-4. Input dialog box

5. With all the required inputs TransCAD will calculate the trip productions for different household groups. The user can output the household trip productions to an existing file or save it as a new file.
6. Once the calculation is done, the user has a choice of continuing to implement another trip purpose. If the user selects "no", the ORP module will aggregate trip productions for each TAP and save the TAP level trip productions into a file. The TAP level trip production results are also connected to the map so that the trip production information is provided in the "info" window by clicking on any TAP.

The trip productions by purpose are input into the disaggregate attraction-end choice model to generate a Production/Attraction matrix. The next section describes the implementation of the attraction-end choice model.

4.3 IMPLEMENTATION OF THE DISAGGREGATE ATTRACTION-END CHOICE MODEL FOR TRIP DISTRIBUTION

4.3.1 Introduction

As has been discussed in Section 4.2, a disaggregate attraction-end choice (DAEC) model has been used in this project to estimate the number of trips produced from and attracted to each zone. Integration of the trip matrices with a GIS-based map display of the region would be helpful in providing the user with a graphical tool to extract trip interchange information. A GIS based visualization of the trip interchanges between zones requires two basic steps: (a) implementation of the computation process in the GIS environment using a macro language, and (b) integration of the trip distribution outputs with a comprehensive map display of the region under consideration. The programming phase is necessary since TransCAD only contains the conventional gravity model for trip distribution and has no current facility for advanced travel demand models.

For the current study, the DAEC model has been programmed using the GIS Development Kit (GISDK) macro language provided by TransCAD. The DAEC macro requires zonal trip productions as input. Therefore, prior to implementation of the trip production phase it is necessary to use the DAEC model. The inputs required for the DAEC macro are summarized in Table 4-1. The next section discusses the requirements for the format and contents of the input files.

Table 4-1. Input file formats and contents for the DAEC macro

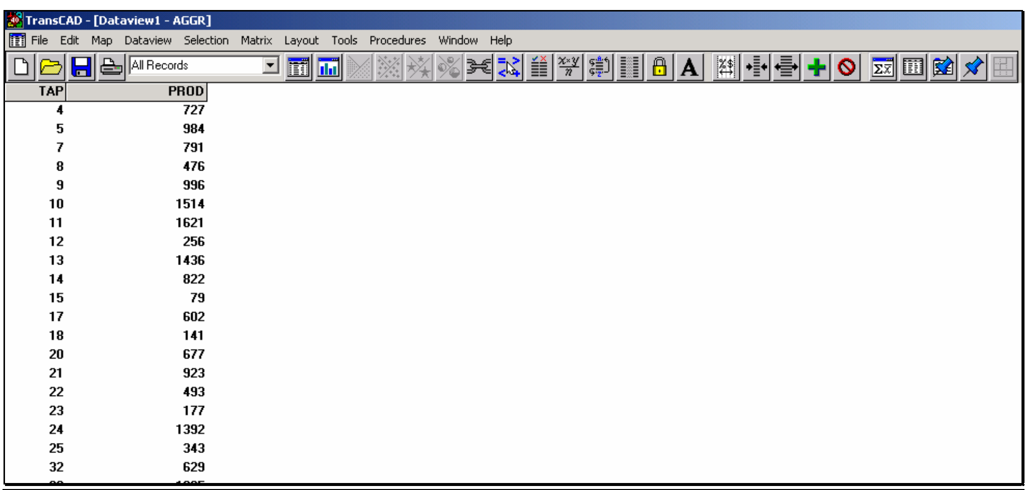
S. No	Inputs	Input File Format	File Fields
1	Trip Production Data (From trip generation phase)	.dbf (DBASE File)	TAPZ, PROD
2	Composite Impedance Matrix	.mtx (TransCAD Matrix)	ROW ID - TAPZ COLUMN ID - TAPZ
3	Socio Demographic Interaction Coefficients (from the DAEC Model)	.dbf (DBASE File)	SDGROUPS, COEFF
4	Land Use Characteristics (from the DAEC Model)	.dbf (DBASE File)	TAPZ, RETAIL, SERVICE, OFFICE, INDUSTRY, INST, TOTEMPL
5	Land Use Coefficients (from the DAEC Model)	Provided by User or .dbf (DBASE File)	LUSE, COEFF

4.3.2 Input File Description

The input files listed in Table 4-1 are prerequisites for the DAEC program. It is necessary that the file formats and contents are the same as indicated in Table 4-1. This section elaborates on the requirements for file structure and contents to use the DAEC macro.

4.3.2.1 Trip Production File

The DAEC macro allows the user to compute trip interchanges from zonal trip productions using traveler socio-demographic and attraction zone characteristics. It is, therefore, necessary to implement the trip production phase before the DAEC macro is used. The trip production file must provide zonal trip production counts. The zones are identified by their TAP numbers and the productions from these zones by a field PROD. In essence, the trip production file must necessarily have two columns: the first column representing the TAP and the second named PROD showing the corresponding trips produced. The trip production file structure is shown in Figure 4-5. This file structure allows the user to compute trip interchanges for each trip purpose separately.



TAP	PROD
4	727
5	984
7	791
8	476
9	996
10	1514
11	1621
12	256
13	1436
14	822
15	79
17	602
18	141
20	677
21	923
22	493
23	177
24	1392
25	343
32	629

Figure 4-5. Trip production file format

4.3.2.2 Impedance Matrix File

An individual's choice of attraction end for a trip is dependent upon the impedance to travel between the production and attraction zones. In this project, a composite impedance matrix has been used for trip distribution. The composite impedance terms capture the combined effect of travel time (both in-vehicle and out-of-vehicle) and cost for each available mode on the utility of choosing a particular attraction zone. The computation of these values has already been discussed in the previous sections. The composite impedance matrix is a square matrix with a size equal to the number of zones in the planning region. The row labels are the production zone ids and the column labels are the attraction zone ids. The number of rows and columns in the impedance matrix should be the same as the number of observations in the trip production file. The matrix file structure and contents are shown in Figure 4-6.

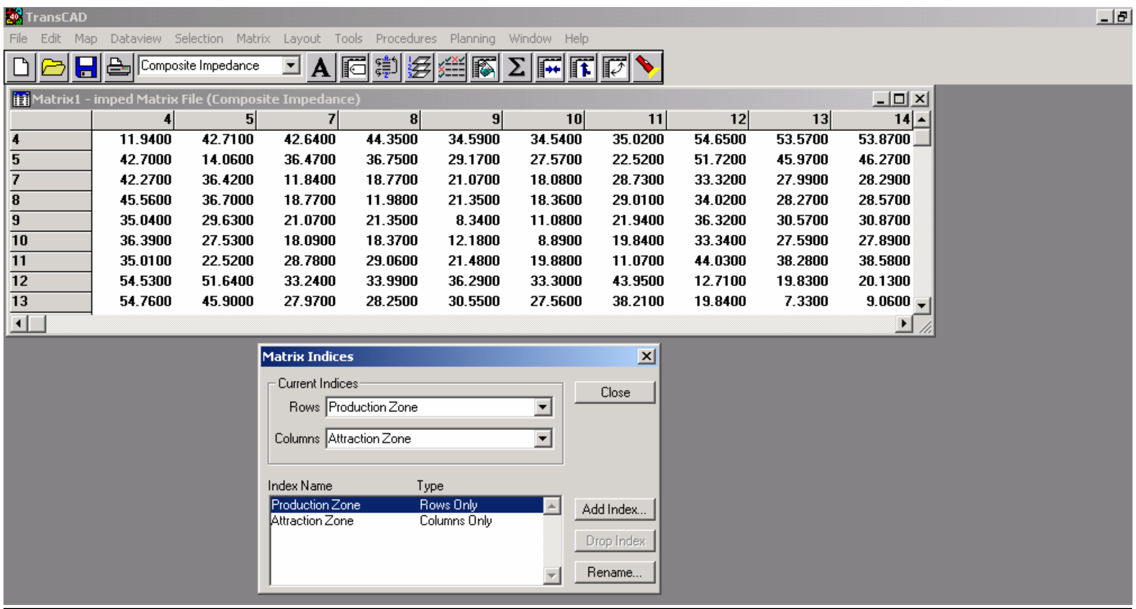
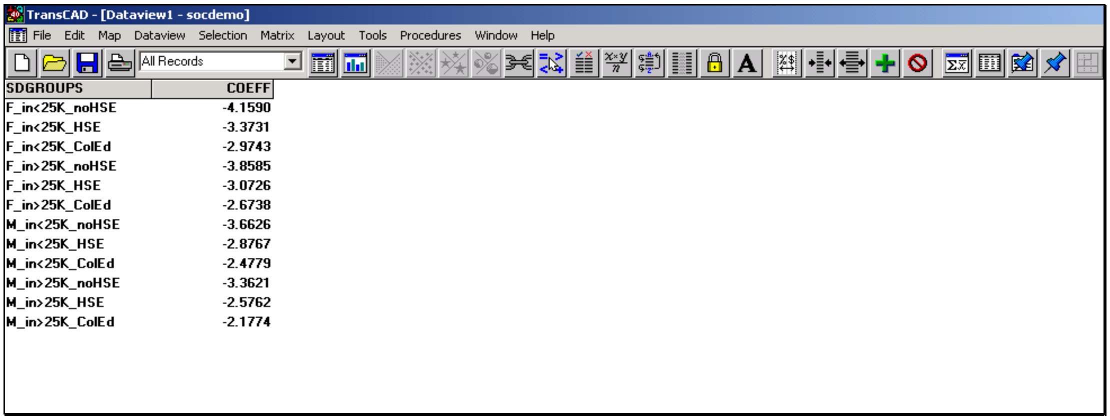


Figure 4-6. Impedance matrix contents

4.3.2.3 Socio-Demographic Coefficient File

People with different socio demographic characteristics have different perceptions of disutility to travel. For this reason, the composite impedance matrix is not the same for all socio-demographic groups. The DAEC model takes this into account through an interaction term between the socio-demographic groups and the composite impedance. The DAEC macro, therefore, seeks a socio-demographic coefficient file. The number of socio-demographic groups varies from one trip purpose to the other. The number of records in the socio-demographic coefficient file is the same as the number of socio demographic groups. The socio-demographic groups are represented by the field SDGROUPS and the corresponding interaction coefficients by the field COEFF. The file structure is shown in Figure 4-7.



SDGROUPS	COEFF
F_in<25K_noHSE	-4.1590
F_in<25K_HSE	-3.3731
F_in<25K_ColEd	-2.9743
F_in>25K_noHSE	-3.8585
F_in>25K_HSE	-3.0726
F_in>25K_ColEd	-2.6738
M_in<25K_noHSE	-3.6626
M_in<25K_HSE	-2.8767
M_in<25K_ColEd	-2.4779
M_in>25K_noHSE	-3.3621
M_in>25K_HSE	-2.5762
M_in>25K_ColEd	-2.1774

Figure 4-7. Socio demographic interaction coefficients

4.3.2.4 Land Use File

The number of trips attracted to a zone also depends on the attraction zone characteristics. The zonal size measures are proxy measures for the number of elemental destinations within a zone. In the current study, total zonal employment has been introduced as a size measure for the HBW purpose, and zonal retail and service employment is used for HBNW and NHB purposes. Zonal office, industrial and institute areas have also been included in the model. The DAEC macro, therefore, also seeks a land use file to compute the trip interchanges between the zones. The land use file must contain information on the zonal retail, service, industrial, institute areas and the total zonal employment. The land use file structure is shown in Figure 4-8.

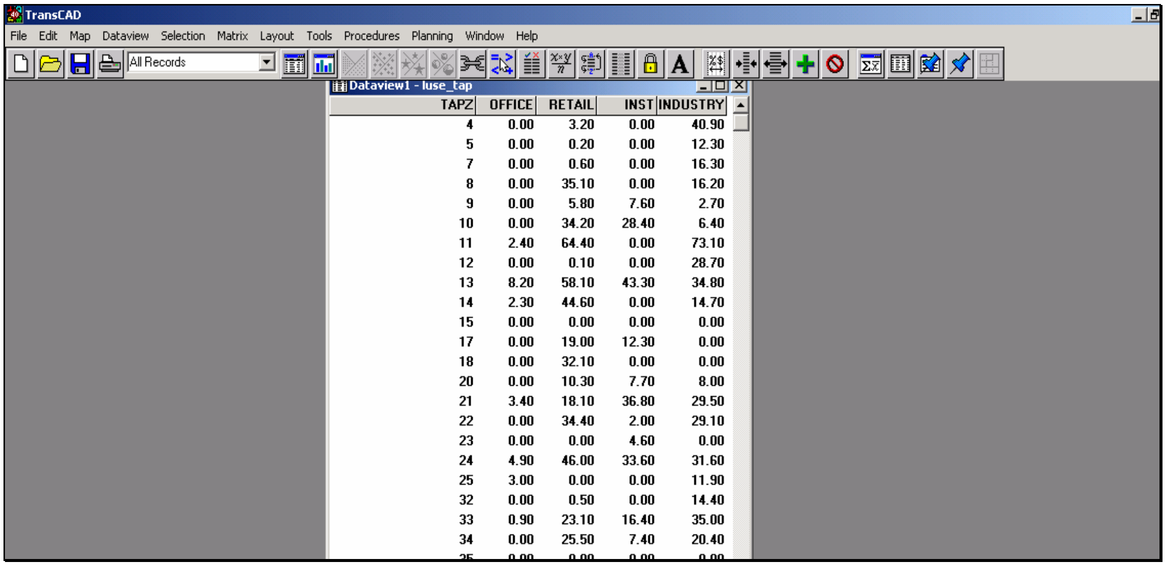


Figure 4-8. The land use file

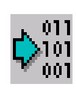
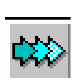
4.3.3 Program Interface and Output File

The DAEC macro guides the user through the input process using dialog boxes and prompt windows. This section describes the input sequence and the program interface.

4.3.3.1 Executing the DAEC Macro

The DAEC macro can be executed using the GISDK Toolkit supported by TransCAD. To open the GISDK toolkit, the user can choose *Tools - Add-ins - GIS Developer's Toolkit* from the TransCAD window. This will open the GISDK toolkit menu bar shown in Figure 4-9.

Figure 4-9. The GISDK toolbox

To compile the DAEC macro, the user must choose the  button in the toolbox and then choose the corresponding resource file that contains the source code. The source code can then be executed by clicking on the  button in the GISDK toolbox. The program will then prompt the user for the name of the macro as shown in Figure 4-10. The actual input process starts after the name of the macro has been entered. For the current project, the DAEC macro is titled *TripDistribution*.

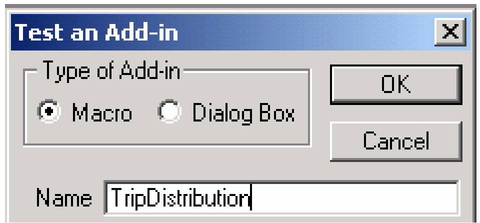


Figure 4-10. Input dialog box for name of macro

4.3.3.2 Program Interface

To start with, the DAEC macro prompts the user for the trip purpose, the zone count and the number of socio-demographic categories relevant for the current trip purpose. The prompt dialog box is shown in Figure 4-11.

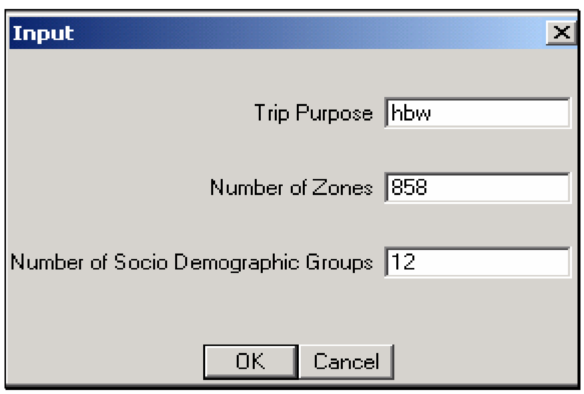


Figure 4-11. Trip purpose and zone count input

Following the trip purpose and zone count inputs, the program prompts the user to choose the composite impedance matrix file. The DAEC macro requires that the impedance matrix file be in the .mtx format supported by TransCAD. The file dialog box is shown in Figure 4-12.

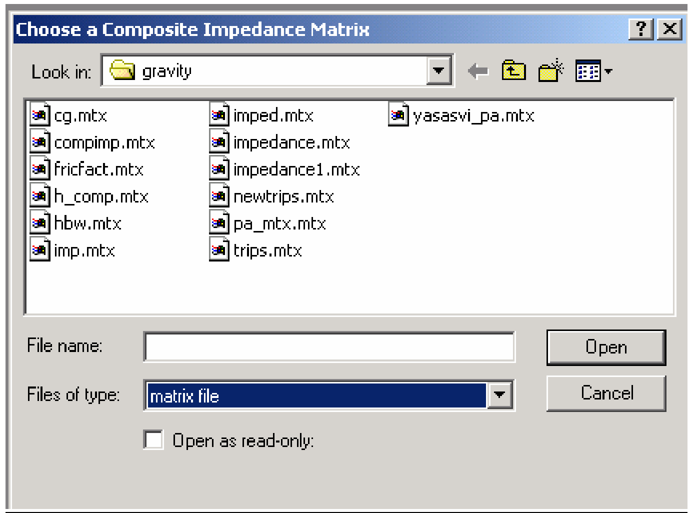


Figure 4-12. Impedance file input

The program then prompts for the trip production file and the socio-demographic coefficients in that order. The files need to be in the .dbf format. Figure 4-13 shows the file dialog boxes.

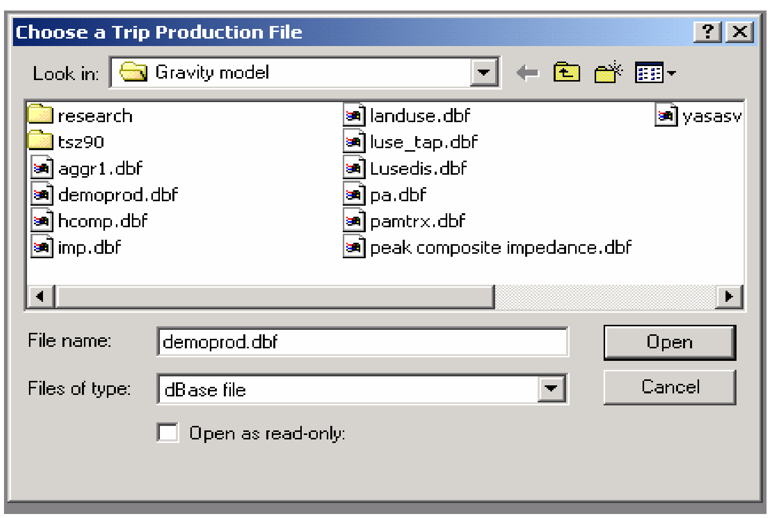


Figure 4-13(a). Input dialog box for trip production file

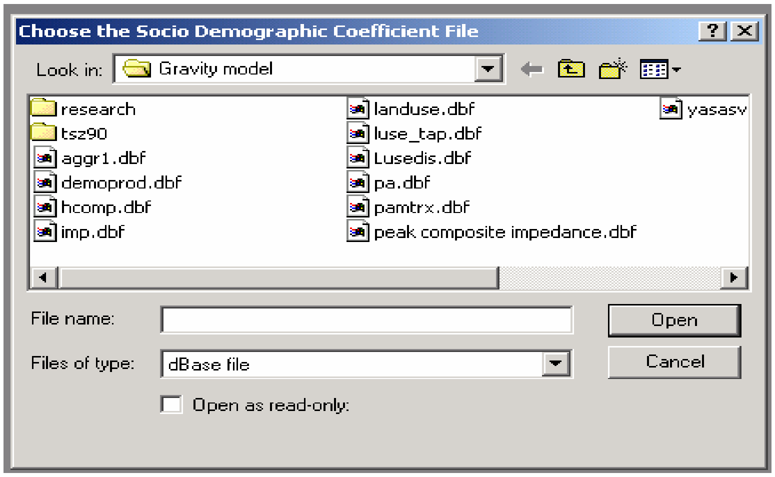


Figure 4-13(b). Input dialog boxes for trip production and socio-demographic coefficients

Figure 4-13. Input Dialog Boxes

Finally, the DAEC macro prompts for the land use file and the land use coefficients. The prompt dialog box and the input dialog box are shown in Figure 4-14.

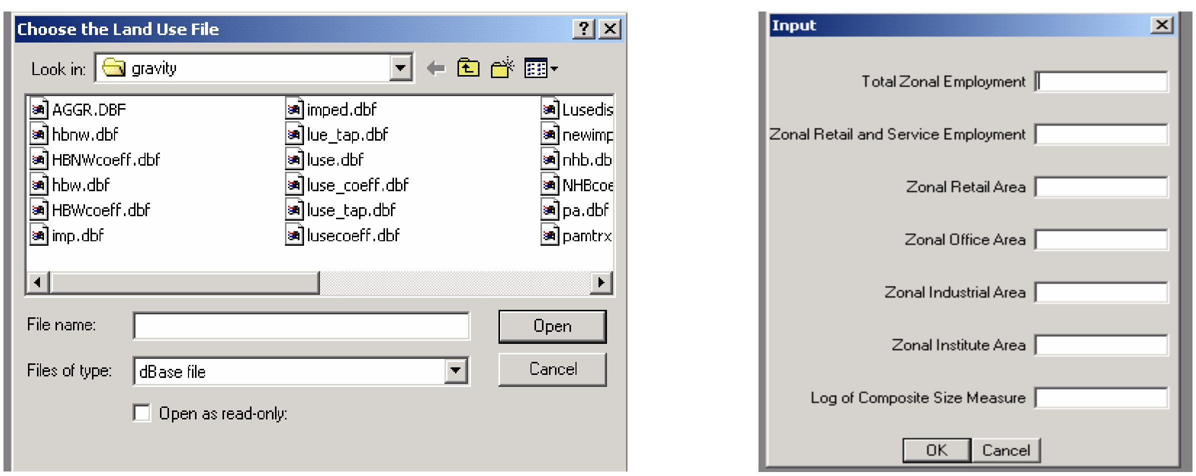
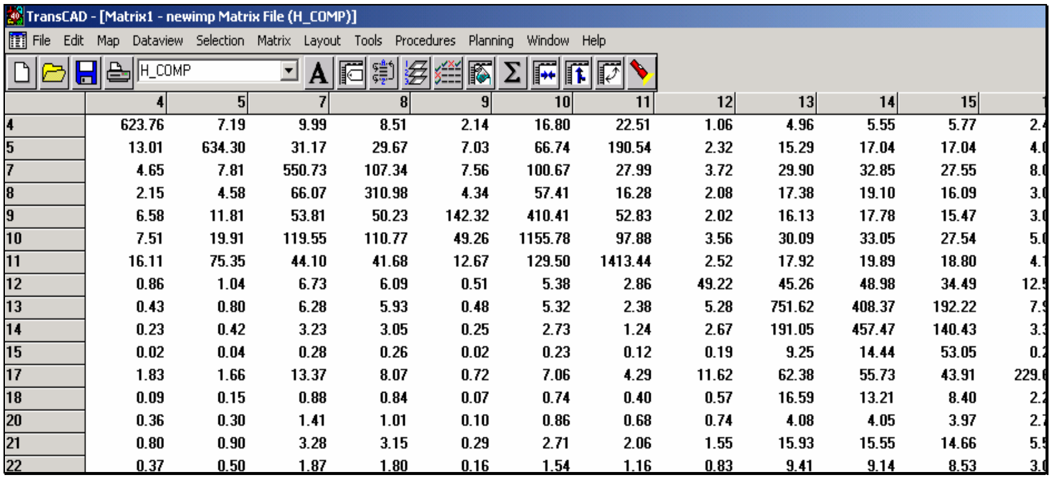


Figure 4-14. Land use file and coefficient inputs

4.3.3.3 The Trip Interchange Matrix Output

The output from the DAEC macro is a matrix file showing the trip interchanges from one zone to the other. The matrix obtained from the trip distribution stage shows the trips produced from a zone and attracted to every other zone. The program informs the user about the directory and file path of the trip matrix using a message dialog box. The trip matrix can then be converted to an O-D Matrix using the "PA to OD" function provided by TransCAD. Figure 4-15 shows a sample trip matrix output file.



	4	5	7	8	9	10	11	12	13	14	15	1
4	623.76	7.19	9.99	8.51	2.14	16.80	22.51	1.06	4.96	5.55	5.77	2.4
5	13.01	634.30	31.17	29.67	7.03	66.74	190.54	2.32	15.29	17.04	17.04	4.0
7	4.65	7.81	550.73	107.34	7.56	100.67	27.99	3.72	29.90	32.85	27.55	8.0
8	2.15	4.58	66.07	310.98	4.34	57.41	16.28	2.08	17.38	19.10	16.09	3.0
9	6.58	11.81	53.81	50.23	142.32	410.41	52.83	2.02	16.13	17.78	15.47	3.0
10	7.51	19.91	119.55	110.77	49.26	1155.78	97.88	3.56	30.09	33.05	27.54	5.0
11	16.11	75.35	44.10	41.68	12.67	129.50	1413.44	2.52	17.92	19.89	18.80	4.1
12	0.86	1.04	6.73	6.09	0.51	5.38	2.86	49.22	45.26	48.98	34.49	12.5
13	0.43	0.80	6.28	5.93	0.48	5.32	2.38	5.28	751.62	408.37	192.22	7.9
14	0.23	0.42	3.23	3.05	0.25	2.73	1.24	2.67	191.05	457.47	140.43	3.3
15	0.02	0.04	0.28	0.26	0.02	0.23	0.12	0.19	9.25	14.44	53.05	0.2
17	1.83	1.66	13.37	8.07	0.72	7.06	4.29	11.62	62.38	55.73	43.91	229.0
18	0.09	0.15	0.88	0.84	0.07	0.74	0.40	0.57	16.59	13.21	8.40	2.2
20	0.36	0.30	1.41	1.01	0.10	0.86	0.68	0.74	4.08	4.05	3.97	2.7
21	0.80	0.90	3.28	3.15	0.29	2.71	2.06	1.55	15.93	15.55	14.66	5.5
22	0.37	0.50	1.87	1.80	0.16	1.54	1.16	0.83	9.41	9.14	8.53	3.0

Figure 4-15. Trip interchange matrix output

4.4 MODE CHOICE AND DEPARTURE TIME CHOICE MODEL IMPLEMENTATION

In this study, multinomial logit model structure is adopted to model mode choice and departure time choice. TransCAD has multinomial logit model implemented for mode split. Our mode choice and departure time choice model were implemented by calling the multinomial logit model evaluation procedure from a resource file that has been compiled into a UI database. The implementation procedure and input data format are entirely consistent with the mode split module in TransCAD. The models also can be directly implemented by using the TransCAD menu. Users may refer to Chapter 7 ("Mode Split and Choice Analysis") in the Travel Demand Modeling manual for more information.

4.5 DISPLAY OF ZONAL TRIP INTERCHANGES

Integration of the trip matrices with a GIS-based map display of the region would be helpful in providing the user with a graphical tool to extract trip interchange information. Therefore, as a sequel to the TransCAD implementation process, the trip interchange matrices have been integrated with a map display of the DFW metropolitan area. The combined information on trip interchanges and geographic features of the region are stored in a TransCAD workspace. For the current project, two separate workspaces were used to store information: (a) a trip production workspace titled *TripProduction.wrk*, for the display of trips produced, and (b) a trip distribution workspace titled *TripDistribution.wrk*, for the display of trip interchanges between zones by mode and departure time. The two workspaces are shown in Figures 4-16(a) and (b).

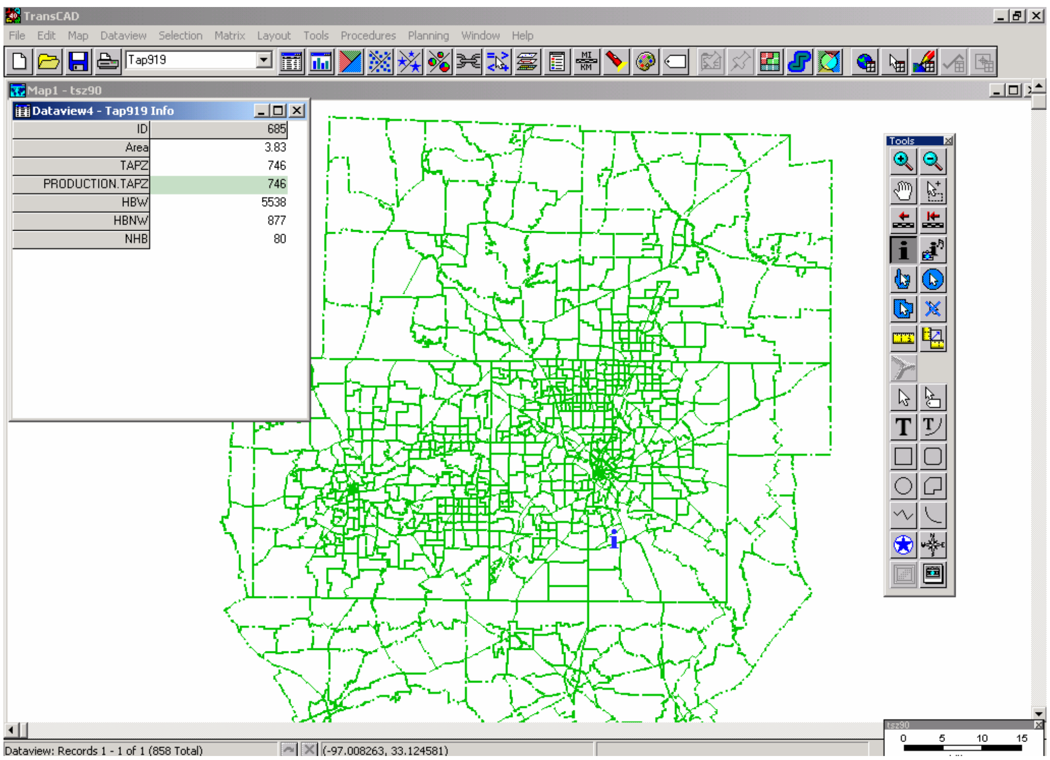


Figure 4-16(a). Workspace for trip production display

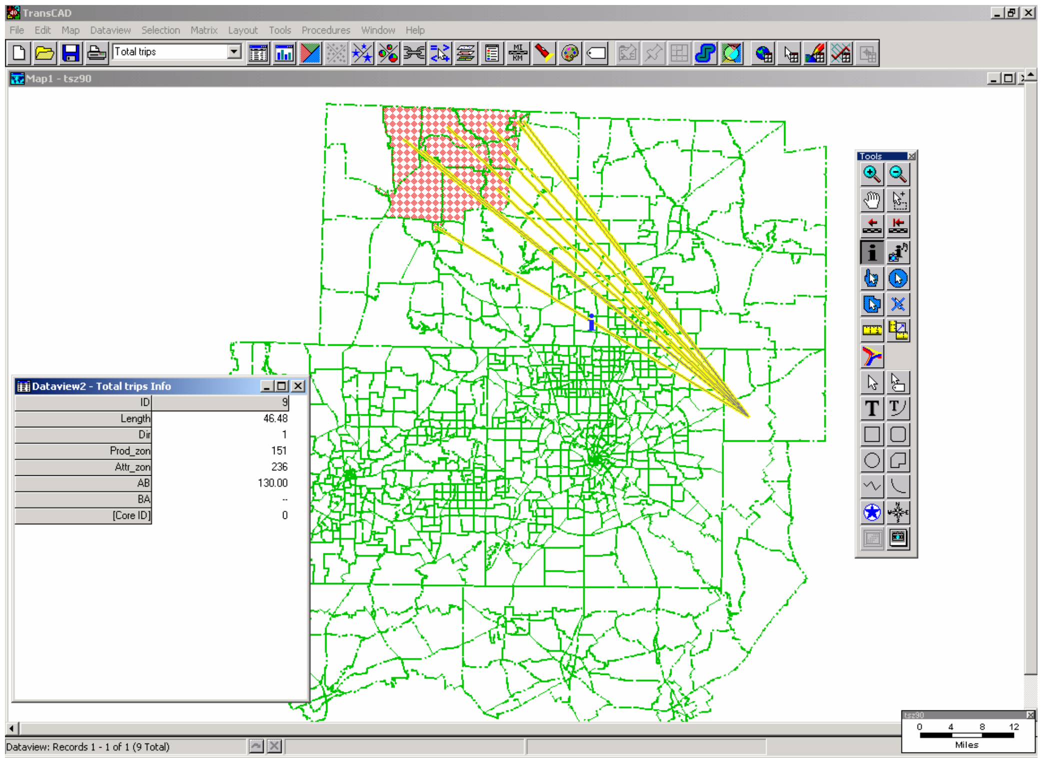


Figure 4-16(b). Workspace for trip interchanges

Figure 4-16. Workspaces for trip productions and interchanges

4.6 IMPLEMENTATION OF MODELS FOR SUPPLEMENTARY TRAFFIC INPUTS TO THE MOBILE EMISSIONS FACTOR MODELS

4.6.1 Introduction

The models developed in this project enhance the accuracy of important inputs to the MOBILE emissions model. For this project several emissions models have been developed and estimated using data for the DFW area (see previous sections). In this section, the implementation of these models within TransCAD is discussed. The implementation focuses on the DFW region and mostly uses data provided by the North Central Texas Council of Governments (NCTCOG).

A couple of comments are in order before continuing. First, the following sections illustrate the implementation of models within TransCAD. In addition to the subsequent descriptions of the implementation, the software files themselves can be found on the compact disk submitted with this report. The *readme* files are also a useful and detailed complement to this discussion. Second, all the models described in this chapter were estimated in an earlier stage of this project. The information necessary for use in implementing the models of supplementary traffic inputs to the MOBILE model is obtained from earlier estimations and relevant input data.

The first model implemented was the VMT mix model. This model calculates the proportion of VMT traveled by each of several vehicle types on each link of the roadway network as a function of link and zonal characteristics.

Another model implemented in this project is the trip time duration model. Calculations for this model are performed at the TAP (Traffic Analysis Process) zonal level. Based on six trip duration "time-bins," this model calculates the fraction of interzonal or intrazonal VMT traveled in each time-bin for each trip purpose/time-of-day (TOD) combination (see Table 4-2). The model also calculates the mean trip time for each purpose/TOD combination in each time bin and the proportion of total trips that each combination comprises. Using this information, the total VMT accrued for each purpose/TOD combination within the TAP zone is predicted. Furthermore, this model can be used to calculate local road VMT (or LRVMT).

Table 4-2. Trip time duration model: levels in each variable category

HB/NHB	Trip Purpose	Time-of-Day	Inter/intrazonal
Non-home	Work	evening	Interzonal
Home	School	morning/peak (am or pm)	Intrazonal
	Social/Recreational	off-peak (am or pm)	
	Shopping		
	Other		

A third model implemented in the GIS framework is the soak time duration model. Soak-time is defined as the amount of time preceding a successful vehicle start that an engine has been turned off. The soak-time duration model is needed to calculate for MOBILE the proportion of trips falling into each of 69 different soak-time intervals. The models developed for soak time analysis are based on the trip purpose preceding the vehicle start, the TOD, and certain zonal attributes (see Table 4-3).

Table 4-3. Soak time duration model and binary logit model used for calculating percentage of first starts: levels in each variable category for each zone

Trip Purpose	Time-of-Day	Inter/intrazonal	MOBILE Time Bins (minutes) 69 total
Home	morning	Interzonal	0
Work	a.m. peak	Intrazonal	0-1,…,29-30
School	a.m. off-peak		30-32,…,58-60
Social/Recreational	p.m. off-peak		60-90,…,690-720
Personal Business	p.m. peak		720+
Shopping	evening		
Other			

4.6.2 Implementation of the VMT Mix Model

4.6.2.1 Embedding the Model in TransCAD

The VMT (vehicle miles traveled) Mix model is used to compute the fractions of total VMT traveled by each vehicle type on each roadway link of the network. The mix of vehicles on the road is an extremely important input for emissions models since different vehicle types emit different amounts of pollutants.

4.6.2.2 Input File Descriptions

Two types of data are needed for predicting VMT mix: roadway link attributes and zonal characteristics. The model is based on the hypothesis that VMT mix on a link is determined by link characteristics (such as speed, number of lanes, and functional class) as well as zonal characteristics, in addition to factors encapsulated in the error term. For instance, an arterial in a heavily industrialized zone probably has more heavy truck travel than an arterial that runs through the central business district (CBD). Therefore, in order to predict VMT mix with this model, the following link and zonal data are needed:

(1) Link data

- functional classification (highway, major/minor arterial, collector/local)
- physical attributes (divided/undivided road, number of lanes)
- free speed variables (< 30 mph, 30-40 mph, 40-55 mph, > 55 mph)
- degree of urbanization (CBD, urban-residential)

(2) Zonal land-use variables

- presence of airport in zone,

- presence of institution(s) (e.g., a church, a school, etc.),
- acreage in office/retail space,
- acreage in manufacturing/warehousing.

4.6.2.3 Program Interface and Outputs

Implementing this model within TransCAD is relatively simple. Once the link and zonal attribute data are collected, only about 10-20 intermediate calculations are needed to arrive at the model results. The model is a multinomial logit model, therefore all of the intermediate calculations are in the easily implementable form of:

$$exp(\beta_i' x_q), \tag{4-1}$$

where β_i' is the vector of model coefficients and x_q is a vector of zonal and link attributes for each link in zone q.

To view the model results, the user needs to open the workspace file VMT_Mix.wrk (see Figure 4-17 below). Note that the chart in the figure does not show all of the variables on the screen at one time; however, by scrolling upward in the chart, the rest of the variables can be viewed. Also note that both the model output as well as the MOBILE counterpart of the model variable were computed (e.g., FR_CAR, the percentage of cars on that link, was converted into LDGV and LDDV). The user can view the VMT mix on any link by using TransCAD's Info tool and clicking on a link on the map.

The dBASE file VMT_PRESENT.DBF (a dBASE file) is another way to use these results. This file is useful for manipulating the data without the map.

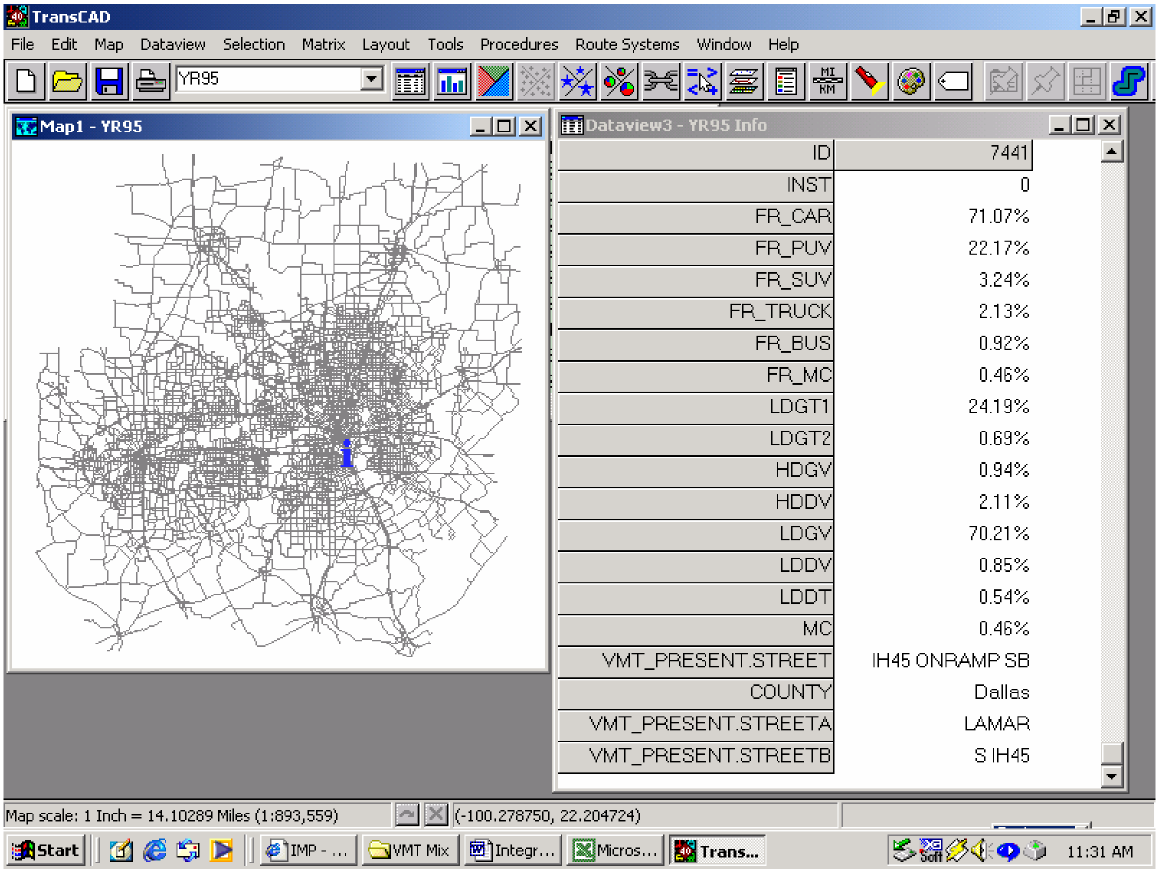


Figure 4-17. Display of the VMT Mix on an IH-45 SB ramp in Dallas, Texas

4.6.3 Implementation of the Trip Time Duration Model

4.6.3.1 Embedding the Model in TransCAD

Implementing the trip time duration model in TransCAD requires much more effort than implementing the VMT Mix model. This is due to the versatility of the trip time duration model – this model can accommodate combinations of ten different trip purposes (five purposes sorted into home-based and non-home-based categories) and six different TOD's. (See Table 4-2.)

4.6.3.2 Input File Descriptions

The following zonal attribute data is needed to implement the model:

- size-related: zonal area, acreage in office space, number of employees in the service sector, manufacturing acreage, retail acreage, and institutional acreage;
- non-size related: household density, median income, and presence of an airport or airport-related infrastructure.

Within TransCAD, the means and variances of the trip time duration distributions were calculated using the parameter estimates obtained in the model estimation stage combined with the data listed above. These means and variances were then used to convert each trip time duration boundary (0 min., 10 min., 20 min., 30 min., 40 min., 50 min., and infinity) into its corresponding normally distributed z-value.

Involving the trip time duration boundaries multiplies the size of the problem. There are six trip time duration bins and seven time-bin boundaries. For each of these boundaries and each trip purpose/TOD/HB-NHB/intrazonal-interzonal combination, the probability distribution function (pdf) and cumulative distribution function (cdf) were calculated (2 calculations for each of the 7 boundaries – the original problem of 120 combinations is therefore multiplied by 14). Ultimately, over 1,500 calculations and normal distribution table lookups were performed in this intermediate stage of the problem.

Within TransCAD, it is easy to calculate the pdf's for these z-values. However, TransCAD contains no functions for evaluating cdf's. Due to this statistical limitation, gathering the cdf-values was a much more intensive effort that required GIS-DK macro programming to obtain cdf-values from a lookup table. Several software programs were therefore written in order to determine the normal cdf's. These files have names such as cdf-a00-a08.rsc, which indicates that the file is a resource file program which uses a lookup table to determine the normal cdf for interzonal trips ("a") for the trip purposes coded 00 through 08.

Next, thousands more formulas were written in order to arrive at the desired FVMT and LRVMT values. The code for these formulas can be found in files with names such as LO.rsc, VMT.rsc, FVMT.rsc, and LRVMT.rsc. The LO.rsc program file is used to calculate the fraction of each trip type in each time-bin, and the average speed in that time-bin. VMT.rsc and FVMT.rsc are used to calculate the total VMT and fraction of VMT, respectively, for each trip type in each time-bin. LRVMT.rsc is used to calculate the VMT accrued on local roads for each trip type.

Note that for FVMT, the fraction of VMT traveled in a certain time-bin, national default values for speed (V^k) from the 1995 NPTS data were used for the calculations in this project (Table 4-4).

Table 4-4. NPTS default average speeds

Time bin	Trip duration	Default speed (mph)
1	0-10 min.	18.96
2	10-20 min.	20.80
3	20-30 min.	26.40
4	30-40 min.	29.14
5	40-50 min.	33.60
6	50+ min.	45.30

Speed for each time bin can also be obtained from local metropolitan area data. Similarly, to calculate local road VMT, it was assumed that the average speed on local roads is 20 mph.
Total VMT on local roads (LRVMT) is:

$$\theta^{l}_{tiz} * v^{l} * I \qquad (4\text{-}2)$$

where v^l is the average speed on local roads (assumed as 20 mph), and I is the total number of intrazonal vehicle trips (obtained from the trip distribution step of the travel demand modeling process).

4.6.3.3 Program Interface and Outputs for the Trip Time Duration Model (Including LRVMT)

This model interface uses the map Tap919.dbd, which is a map of the DFW region divided into TAP zones, and the dataview TripDFinal-base.dvw, which contains the original variables, the embedded formulas and the model results. Clicking on the map with the Info tool displays the model results for that zone. Both can be opened separately, or together by opening the workspace file FVMT.wrk. For trips in each time-bin sorted by trip purpose and time period, it also contains

1. the calculated proportions of those trips;
2. the mean trip duration of those trips (per time period and purpose);
3. the total VMT (in miles, and accumulated across all time bins) accrued by each of those types of trips; and
4. the fraction of VMT (fvmt) accrued by each of those types of trips.

Figure 4-18 below shows the resulting TransCAD display. VMT** indicates (3.) above, and FVMT** indicates (4.) above.

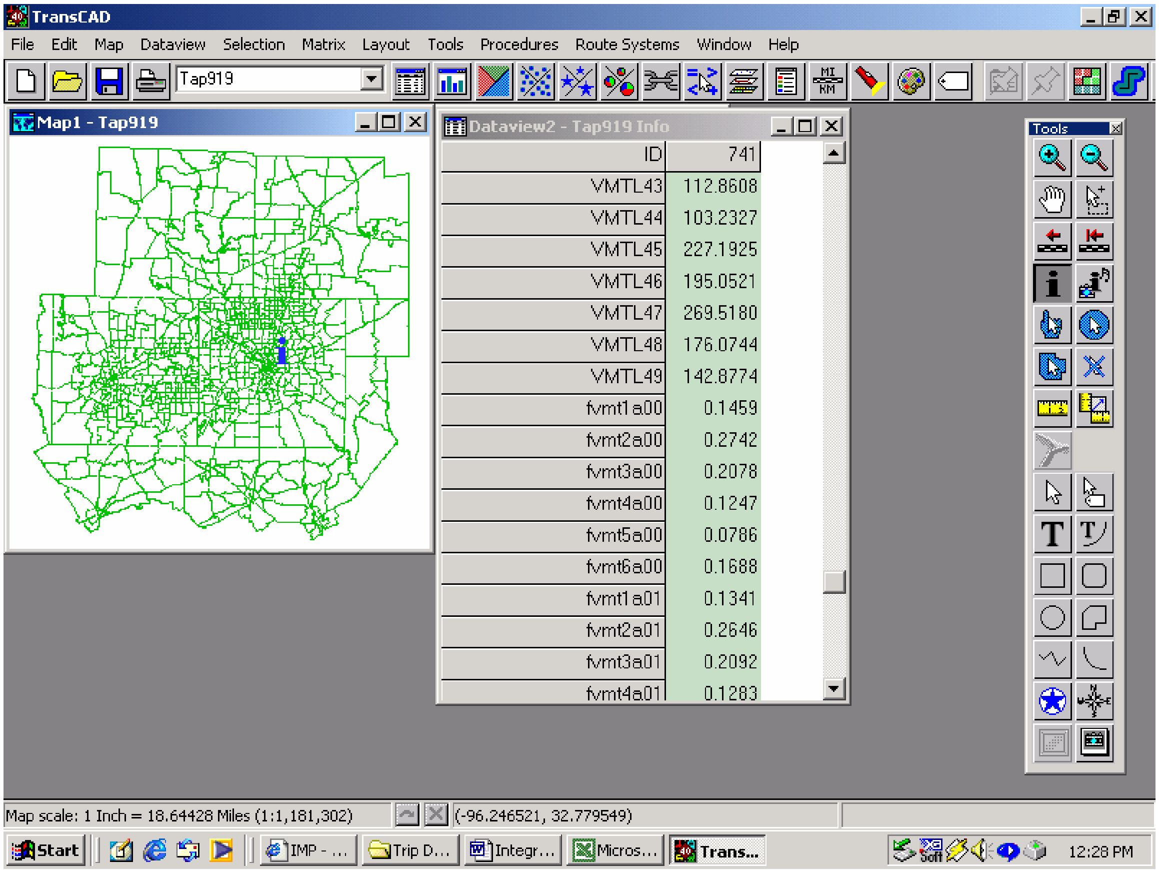


Figure 4-18. Some DFW trip time duration model results

Similarly, VMT accrued on local roads (LRVMT) can be viewed by opening the workspace LRVMT.wrk or by opening the dataview TOTAL_LRVMT.dvw (see Figure 4-19).

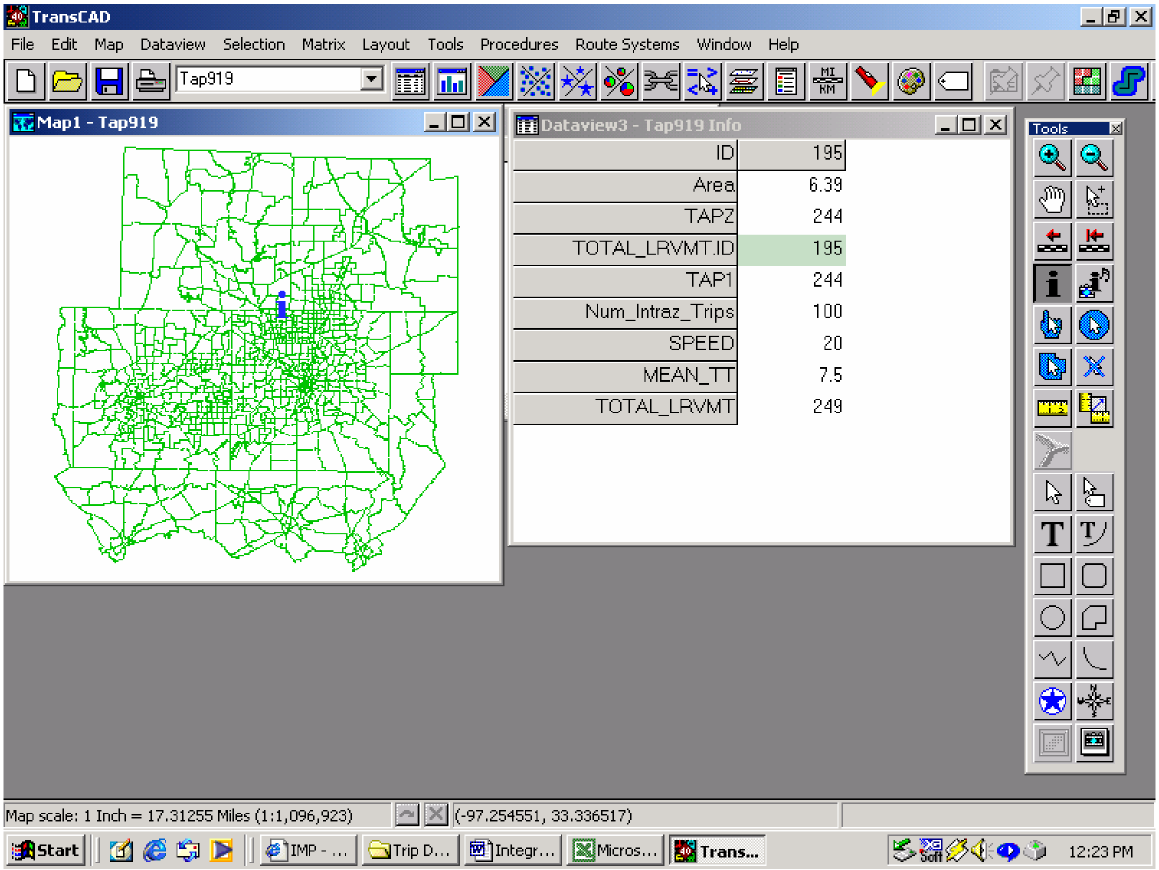


Figure 4-19. Local road VMT results

4.6.4 Implementation of the Soak-Time Duration Model

4.6.4.1 Embedding the Model in TransCAD

Implementing the Soak Time Duration model in TransCAD requires significant effort. Due to the number of MOBILE soak time bins (69), combined with the number of trip purpose/TOD/interzonal-intrazonal combinations, over 5,000,000 cdf-values must be calculated in order to find the soak-time distribution: approximately 5,000 cdf's for each of the 858 zones (see Table 4-3). Given TransCAD's limited statistical capabilities (i.e., it cannot calculate the normal cdf), the statistical portion of the problem was done using the statistical software SPSS.

4.6.4.2 Input File Descriptions

The same input files were used to implement the soak time duration model as the files used to implement the trip time duration model. For demonstration purposes, however, only one trip type was implemented for the soak time duration model. (This is all explained in greater detail in the *readme* files, along with an outline of suggested code to write to expedite the process of implementing the rest of it.)

Using the data and the model parameters, the fraction of soaks in time-bin *k* across first and non-first starts for zone *z*, time period *t*, and activity purpose *i* was computed. First, the normal cdf equations were written in Microsoft Excel, mostly using the Copy, Paste, and Merge Cells commands. Using a spreadsheet for this process greatly reduces the amount of time required to create the equations. Second, the equations were saved in *.txt format, then reopened in Microsoft Word where it was easy to replace or remove certain characters (in this case, quotation marks) that resulted from the transition from spreadsheet to word processor format. Third, the equations were copied and pasted into an SPSS syntax file as computation commands. Running these commands created a 35 MB file in *.sav format.

4.6.4.3 Program Interface and Outputs for the Soak Time Duration Model

Importing the resulting 35-MB SPSS file into TransCAD proved to be extremely difficult, so for the purposes of illustration the subsequent material is based on an excerpt that contains the distribution of interzonal first-start soak-times sorted by time-bins for pairing "10": morning (1) and home purpose (0). The lognormal mean and standard deviation for this pairing are 6.34 and 0.22, respectively; so, for interzonal first-start trips in this TAPZ occurring during the morning period for which the trip purpose preceding the vehicle start was the "home purpose," the mean soak-time is 567 minutes (which is in time-bin #62). (See Figure 4-20.)

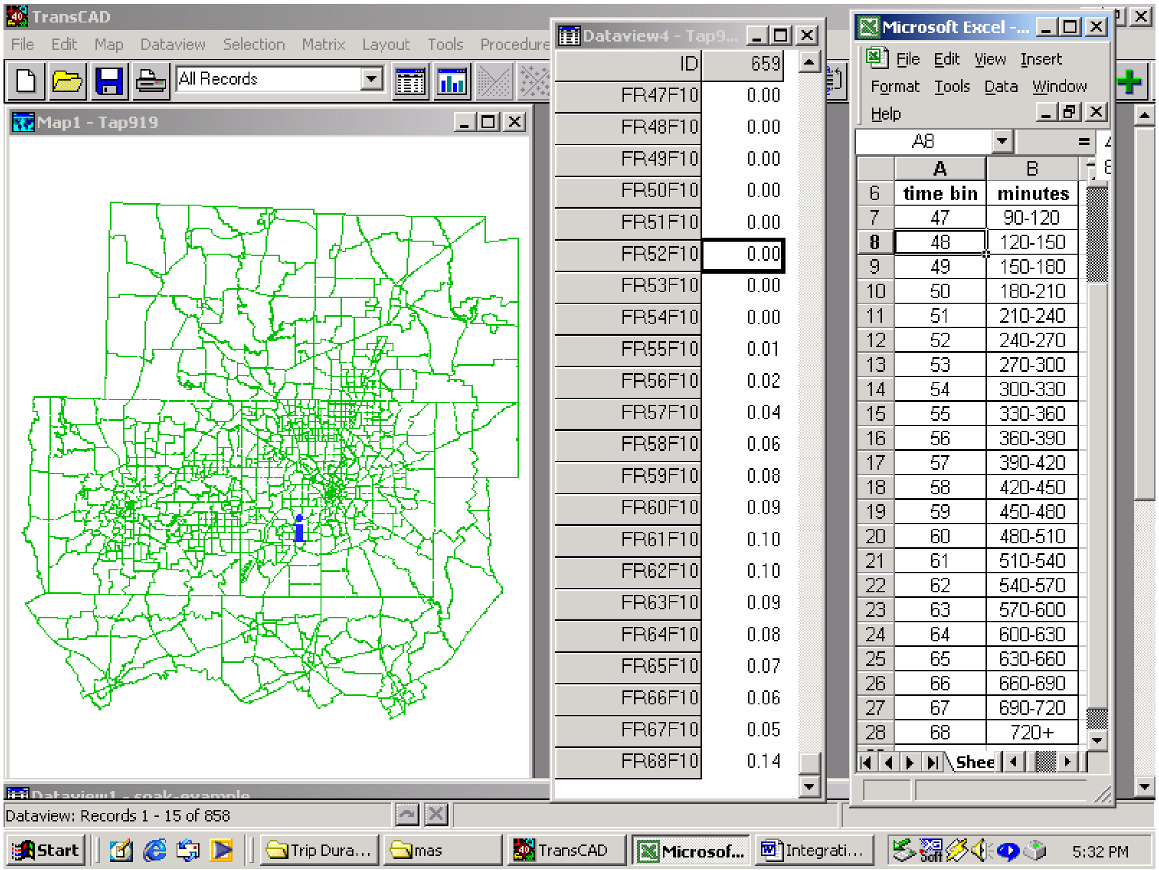


Figure 4-20. Soak time duration distribution for morning and home purpose

4.6.5 General Notes on Implementation Challenges

The main difficulty faced when implementing these models within TransCAD was due to the versatility of the models and the computational limitations of TransCAD. For models like these that can accommodate so many trip purposes, TODs, time-bins, etc., the number of calculations to perform becomes enormous. However, the problem size is challenging mainly because of the relative lack of statistical functions in TransCAD. It is generally very time-consuming to build all the necessary formulas into TransCAD using tables and dataviews. For the non-statistical calculations, it may be easier to implement the models using matrices in TransCAD (if possible). However, practitioners and academics both will probably find it fastest and easiest to perform the computations in a spreadsheet or statistical program, then import the relevant results back into TransCAD.

CHAPTER FIVE – SUMMARY AND RECOMMENDATIONS

Transportation planners, traditionally used to focusing on regional travel forecasting, now have the added responsibility of providing traffic data to state air quality agencies for use in mobile source emissions analysis. Coordinated efforts among land use planners, travel demand planners, and air quality planners are needed to ensure the provision of safe and efficient transportation systems while also addressing environmental concerns. This is particularly so because mobile source emissions constitute a major fraction of total atmospheric emissions. Consequently, many metropolitan areas and states are depending on transportation control measures (TCMs) to reduce mobile source emissions as part of an overall strategy to reduce atmospheric emissions. Within this context, it is important to develop a procedure to determine which TCMs (or combination of TCMs) have the most beneficial impact in terms of mobility, emissions, and cost. It is necessary that such a procedure be methodologically sound yet application friendly, and be capable of analyzing the effects of demographic changes and transport policy actions. The current research contributes toward the development of such an integrated transportation air quality procedure. Specifically, the research has: (a) Refined the structure and specification of travel models, (b) Developed models for supplementary traffic inputs, including VMT mix travel time duration, local VMT and soak-time duration, and (c) Integrated all models within a Geographic Information System (GIS) architecture.

5.1 FINDINGS

There are four broad findings from this research, which are discussed in turn in the subsequent paragraphs.

First, sociodemographic and employment related attributes have a significant impact on travel-related decisions of households and individuals. While this result is not surprising, the state of the practice in travel demand modeling continues to use very limited specifications of sociodemographic variables in the modeling process. Such limited demographic variable specifications can, and in general will, lead to misinformed transportation planning and air quality decisions because of projected changes in demographic and employment-related trends over the next few decades (including aging of the population, a decrease in the number of households with children, and more employed individuals). In addition, our results indicate differential sensitivities of sociodemographic groups to transportation system performance. Accommodating such differential travel sensitivities of demographic groups is important for accurate evaluation of transportation control measure as well as for environmental justice considerations in transport policy analysis.

Second, it is practical and very feasible to apply models in forecasting mode even when they involve several explanatory demographic variables. One of the reasons often provided for the use of a limited specification of demographic variables in estimation is that it renders the forecasting process manageable. However, this research project has developed and applied a forecasting methodology that can be integrated within a GIS-based platform, even if several explanatory variables are used during estimation. The methodology, which is rather simple, entails the generation of a synthetic population of households and individuals based on current or projected aggregate population demographic distributions.

Third, current approaches to VMT mix, local road VMT, soak time distribution, and VMT by trip time duration can be substantially improved by developing models based on local vehicle classification counts and survey data. Since the emissions computations in the MOBILE model are very sensitive to these inputs, it is important that metropolitan planning organizations consider pursuing such efforts.

Fourth, visualization of the intermediate and final traffic output results graphically on the Texas network aids in understanding traffic patterns and provides an effective intuitive means to checking model functionality.

5.2 RECOMMENDATIONS

Our recommendations are provided under two categories: Implementation and Further Research.

Implementation Recommendations

The models developed as part of the integrated transportation air quality procedure can be implemented for the North Central Texas Council of Governments (NCTCOG) area. The models for travel demand will need additional effort to ensure compatibility with the current framework adopted by NCTCOG; however, the models for supplementary traffic inputs can be used as is by

NCTCOG for planning purposes. Implementation of the travel demand and supplementary traffic model formulations for other metropolitan areas will require model estimations based on data collected locally. It is recommended that TxDOT pursue such implementation-related work for other Texas metropolitan areas.

Research Recommendations

The U.S. Environmental Protection Agency (EPA) has now developed the MOBILE6 emissions factor model. The structure as well as the input needs for MOBILE6 are different from those of its predecessor MOBILE5. The changes in MOBILE6 are valuable and should result in more accurate mobile source emissions estimates. At the same time, the MOBILE6 model offers substantial opportunities and poses important challenges to improve traffic inputs. Among these inputs are (a) fleet characterization data (projections of future vehicle fleet size, and fraction of travel by a multidimensional breakdown based on vehicle age, mileage accumulation rate, and thirty vehicle types), (b) separate traffic-related variables for weekdays and weekends in emissions modeling, and (c) a very high temporal resolution during the day for all traffic indicators. It is recommended that TxDOT place a high priority on the development of models that are capable of accurately providing such traffic inputs to the MOBILE6 model.

REFERENCES

Abkowitz, Mark D. (1981) An Analysis of the Commuter Departure Time Decision. *Transportation*, Vol. 10, pp. 283 - 297.

Agyemang-Duah, K., Anderson, W.P. and Hall, F.L. (1995) Trips Generation for Shopping Travel, *Transportation Research Record*, 1493, pp12-20.

Agyemang-Duah, K. and Hall, F.L. (1997) Spatial Transferability of an Ordered Response Model of Trip Generation, *Transportation Research*, Vol. 31A, No. 6, pp389-402.

Allen, W.G., and Davies, G.W. (1993) A New Method for Estimating Cold-Start VMT. *Compendium of Technical Reports from ITE 63rd Annual Meeting*, The Hague, The Netherlands, pp. 224-229.

Ben-Akiva, M. and Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge.

Bhat, C.R. (1997) Work Travel Mode Choice and Number of Nonwork Commute Stops, *Transportation Research*, Vol. 31B, No. 1, pp41-54.

Bhat, Chandra R. (1998a) Analysis of Travel Mode and Departure Time Choice for Urban Shopping Trips. *Transportation Research* B, Vol. 32, No. 6, pp. 361-371.

Bhat, Chandra R. (1998b) Accommodating Flexible Substitution Patterns in Multi-Dimensional Choice Modeling: Formulation and Application to Travel Mode and Departure Time Choice. *Transportation Research* B, Vol. 32, No. 7, pp. 455-466.

Bhat, Chandra R. (1998c) A Model of Post Home-Arrival Activity Participation Behavior. *Transportation Research* B, Vol. 32, No. 6, pp. 387-400.

Bhat, C.R. and Pulugurta, V. (1998d) A Comparison of Two Alternative Behavioral Choice Mechanisms For Household Auto Ownership Decisions, *Transportation Research*, Vol. 32B, No.1, pp. 61-75.

Bhat, Chandra R. (1999a) Duration Modeling. Forthcoming in *Handbook of Transport*, Kenneth Button and David Hensher, eds.

Bhat, Chandra R. (1999b) A Multi-Level Cross-Classified Model for Discrete Response Variables. Forthcoming in *Transportation Research*.

Bhat, C. R., and R. Misra (1999c) Discretionary Activity Time Allocation of Individuals between In-Home and Out-of-Home and between Weekdays and Weekends. *Transportation*, Vol. 26, No.2, pp. 193–209.

Bhat, C. R., and Koppelman, F. S. (1999d) A Retrospective and Prospective Survey of Time-Use Research. *Transportation*, Vol. 26, No. 2, pp. 119-139.

Bhat, C. R. (2000) Modeling the Commute Activity-Travel Pattern of Workers: Formulation and Empirical Analysis. Technical Paper, Department of Civil Engineering, The University of Texas at Austin.

Brodtmen, K. J. and Fuce, T. A. (1984) Determination of Hot and Cold Start Percentages in New Jersey. Report FHWA/NJ-84/001, New Jersey Department of Transportation.

Cambridge Systematics, Inc. (1994) Short-Term Travel Model Improvements, Final Report. Prepared for the U.S. Department of Transportation and the U.S. Environmental Protection Agency.

Chatterjee, A., Reddy, P. M., Venigalla, M., and Miller, T. L. (1996) Operating Modes Fractions on Urban Roads Derived by Traffic Assignment. *Transportation Research Record* Vol.1520, pp. 97-103.

Chatterjee, A. et al. (1997) Improving Transportation Data for Mobile Source Emission Estimates. National Cooperative Highway Research Program Report 394, TRB, National Research Council, Washington, D.C.

Chin, Anthony T. H. (1990) Influences on Commuter Trip Departure Time Decisions in Singapore. *Transportation Research A*, Vol. 24, No. 5, pp. 321-333.

Davis, S. C. (1997) *Transportation Energy Data Handbook.* Prepared by the Oak Ridge National Laboratory for the Office of Transportation Technologies, U.S. Department of Energy.

Deakin, Harvey, Skabardonis, Inc. (1993) *Manual of Regional Transportation Modeling Practice for Air Quality Analysis*. The National Association of Regional Councils, Washington, D.C.

Ellis, G. W., Camps, W. T., and Treadway, A. (1978) The Determination of Vehicular Cold and Hot Operating Mode Fractions for Estimating Highway Emissions. State of Alabama Highway Department.

Fotheringham, A. (1983) Some Theoretical Aspects of Destination Choice and Their Relevance to Production-Constrained Gravity Models. *Environment and Planning*, Vol. 15A, pp. 1121-1132.

Frank, L.D., Stone, Jr., B., and Bachman, W. (2000) Linking Land-Use with Household Vehicle Emissions in the Central Puget Sound: Methodological Framework and Findings. *Transportation Research*, Part D, Vol. 5, pp. 173-196.

Glover, E. L., and Brzezinski, D. J. (1998) Trip Length Activity Factors for Running Loss and Exhaust Running Emissions. Report Number M6.FLT.005, U.S. Environmental Protection Agency Assessment and Modeling Division.

Gordon, Peter, Ajay Kumar, and Harry W. Richardson, (1988) Beyond the Journey to Work. *Transportation Research A*, Vol. 22, No. 6, pp. 419-426.

Goulias, K. G., Pendyala, R. M. and Kitamura, R. (1988) Practical Method for the Estimation of Trip Generation and Trip Chaining. *Transportation Research Record*, Vol. 1285, pp. 47-56.

Gourieroux, C., A. Monfort, and A. Trognon (1984) Pseudo-Maximum Likelihood Methods: Theory. *Econometrica,* Vol. 52, pp. 681–700.

Greene, W.H. (1997) *Econometric Analysis*. Prentice-Hall International, pp. 948-952.

Hamed, Mohammed M., and Fred L. Mannering. (1993) Modeling Travelers' Postwork Activity Involvement: Toward a New Methodology. *Transportation Science*, Vol. 27, No. 4, pp. 381-394.

Hendrickson, Chris, and Edward Plank. (1984) The Flexibility of Departure Times for Work Trips. *Transportation Research* A, Vol. 18, No. 1, pp. 25-36.

Hensher, David A., and Fred L. Mannering. (1994) Hazard-Based Duration Models and their Application to Transport Analysis. *Transport Reviews*, Vol. 14, No. 1, pp. 63-82.

Hunt, J. D., and D. M. Patterson. (1996) A Stated Preference Examination of Time of Travel Choice for a Recreational Trip. *Journal of Advanced Transportation*, Vol. 30, No. 3, pp. 17-44.

Johnson, N., and Kotz, S. (1970) *Distributions in Statistics – Continuous Univariate Distributions*. John Wiley & Sons.

Johnston, J., and DiNardo, J. (1997) *Econometric Methods*, fourth edition, The McGraw-Hill Companies, Inc.

Kanafani, A. (1983) *Transportation Demand Analysis*. The McGraw-Hill Companies, Inc.

Koppelman, Frank, Chandra Bhat, Vaneet Sethi, and Bruce Williams. (1999) A Self-Instructing Manual on Discrete Choice Modeling, Draft Final Report Submitted to the U.S. Department of Transportation.

Kumar, Ajay, and David Levinson, (1995) Temporal Variations on Allocation of Time. In *Transportation Research Record,* Vol. 1493, TRB, National Research Council, Washington, D.C., pp. 118-127.

Kurani, K.S., and Lee-Gosselin, M.E.H. (1996) Synthesis of Past Activity Analysis Applications. Activity Based Travel Forecasting Conference, New Orleans, Louisiana.

Levinson, David, and Ajay Kumar. (1993) Integrating Feedback Into Transportation Planning Model: Structure and Application. In *Transportation Research Record,* Vol. 1413, TRB, National Research Council, Washington, D.C., pp. 70-77.

Maddala, G. S. (1983) *Limited Dependent and Qualitative Variables in Econometrics,* Cambridge University Press, Cambridge.

Mannering, Fred L. (1989) Poisson Analysis of Commuter Flexibility in Changing Routes and Departure Times. *Transportation Research* B, Vol. 23, No. 1, pp. 53-60.

Mannering, Fred L., and Mohammed M. Hamed. (1990) Occurrence, Frequency, and Duration of Commuters' Work-to-Home Departure Delay. *Transportation Research* B, Vol. 24, No. 2, pp. 99-109.

McCafferty, Desmond, and Fred L. Hall (1982) The Use of Multinomial Logit Analysis to Model the Choice of Time to Travel*.* *Economic Geography*, Vol. 36, No. 3, pp. 236-246.

McFadden, D. (1978) Modeling the Choice of Residential Location. *Spatial Interaction Theory and Planning Models*, edited by Karlquist, A., Lundquist, L., Snickbars, F., and Weibull, J.W. North Holland, Amsterdam, pp75-96.

Mckelvey, R. D. and Zavoina, W. (1975) A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology*, Vol. 4, No.2, pp. 103-120.

Meurs, H. (1989) Dynamic Analysis of Trip Generation. Paper presented at the International Conference on Dynamic Travel Behavior Analysis, Kyoto, Japan.

Monzon, J., Goulias, K. and Kitamura, R. (1989) Trip Generation Models for Infrequent Trips, *Transportation Research Record*, 1220, pp. 40-46.

Nair, H.S., Bhat, C.R., and Kelly, R.J. (2000) Modeling Soak-time Distribution of Trips for Mobile Source Emissions Forecasting: Techniques and Applications. Working Report, The University of Texas at Austin.

National Cooperative Highway Research Program (NCHRP) *Research Results Digest* (1998) Development of an Improved Framework for the Analysis of Air Quality and Other Benefits and Costs of Transportation Control Measures. Transportation Research Board, National Research Council.

Office of Mobile Sources, U.S. Environmental Protection Agency (1993) Federal Test Procedure Review Project: Preliminary Technical Report. EPA 420-R-93-007. http://www.epa.gov/oms/regs/ld-hwy/ftp-rev/ftp-summ.txt.

Ortuzar, Juan de Dios, and Luis G. Willumsen. (1990) *Modelling Transport*. Chichester, West Sussex, England: Wiley.

Papke, L. E., and J. M. Wooldridge (1996) Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates. *Journal of Applied Econometrics*, Vol. 11, pp. 619–632.

Pas, Eric I. (1995) The Urban Transportation Planning Process. In *The Geography of Urban Transportation*. S. Hanson, ed., New York: Guilford Press.

Small, Kenneth A. (1982) The Scheduling of Consumer Activities: Work Trips. *The American Economic Review*, Vol. 72, No. 3, pp. 467-479.

Small, Kenneth A. (1997) A Discrete Choice Model for Ordered Alternatives. *Econometrica*, Vol. 55, No. 2, pp. 409-424.

U.S. Census Bureau. (1999a) Interracial Tables. Population Division, Racial Statistics Branch. http://www.census.gov/population/www/socdemo/interrace.html.

U.S. Census Bureau. (1999b) Population Projections. Population Division, Population Projections Branch. http://www.census.gov/population/www/projections/popproj.html.

U.S. Census Bureau. (1999c) Projected Number of Families with Children under 18 Years by Type: 1995 to 2010. Population Division, Population Projections Branch. http://www.census.gov/population/www/projections/nathh.html.

U.S. Census Bureau. (1999d) Working at Home. Population Division, Journey to Work and Migration Statistics Branch. http://www.census.gov/population/www/socdemo/workathome/wkhtab.html.

Venigalla, M., Chatterjee, A., and Bronzini, M.S. (1999) A Specialized Equilibrium Assignment Algorithm for Air Quality Modeling. *Transportation Research,* Part D, Vol. 4, pp. 29-44.

Wooldridge, J. M. (1991) Specification Testing and Quasi-Maximum Likelihood Estimation," *Journal of Econometrics*, Vol. 48, pp. 29–55.