EMPIRICAL MEASUREMENTS OF TRAVELERS' VALUE OF TRAVEL TIME RELIABILITY

A Thesis

by

SANTOSH RAO DANDA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Mark Burris
Committee Members,	Clifford Spiegelman
	Bruce Wang
Interim Head of Department,	Robin Autenrieth

August 2014

Major Subject: Civil Engineering

ABSTRACT

Travel time and travel time reliability are two fundamental factors influencing travel behavior and demand. The concept of the value of time (VOT) has been extensively studied, and estimates of VOT have been obtained from surveys and empirical data. On the other hand, although the importance of value of reliability (VOR) is appreciated, research related to VOR is still in its early stages. VORs have been estimated using surveys but has almost never estimated using empirical data.

This research used empirical data to take an initial step toward understanding the importance of travel time reliability. Katy Freeway travelers face a daily choice between reliable tolled lanes and less reliable but untolled lanes. An extensive dataset of Katy Freeway travel was used to examine the influence of time, reliability, and toll on lane-choice behavior. Lack of clarity on how travelers' perceive travel time reliability meant different measures of reliability had to be tested to see which best represents travelers' perception. In this research, three different measures of reliability were used, namely, standard deviation of travel time, coefficient of variation of travel time and travel time standard deviation relative to the total trip time.

Lane choice was estimated using multinomial logit models. Basic models, including only travel time and toll, yielded reasonable results. Models included VOTs of $1.53/hour, $6.05/hour, and $9.05/hour for off-peak, shoulder, and peak-period travelers, respectively. However, Adding different measures of reliability like standard deviation and coefficient of variation to the models resulted in counter-intuitive results. Positive

coefficients for unreliability of travel time were obtained indicating that travelers, at least on the Katy Freeway, do not value travel time reliability as has been theorized in earlier studies on the same. It was concluded that additional research on how travelers perceive the reliability and time savings on MLs is needed because modeling real-world choices of MLs using empirical data and the standard definitions of reliability and time savings did not concur with the existing theory on travel time reliability and led to counter-intuitive results.

ACKNOWLEDGEMENTS

This research would not have been possible without the support and guidance of my advisor and committee chair, Dr. Mark Burris. I thank him for having faith in me and entrusting me with this research assignment towards the beginning of my masters' program. His invaluable inputs, sharp understanding and clarity of thought guided me through the road blocks of this thesis. Above all, I thank him for his patience and for always keeping his door open for questions and discussions.

I would also like to thank Dr. Clifford Spiegelman, Dr. Bruce Wang and Dr. Dominique Lord for serving on my thesis committee. I am grateful for their valuable feedback that helped look at the research problem from multiple different perspectives and ensure the integrity of this research.

The thesis was, in part, an outcome of the research that was performed under a grant from the U.S. Department of Transportation University Transportation Centers Program for the Southwest University Transportation Center, which is funded, in part, with general revenue funds from the State of Texas and funding from the Texas Department of Transportation (TxDOT). I am grateful for their support. I would also like to thank Xiaozheng (Sean) He for his insights and guidance on this project. I also appreciate the sharing of data by TxDOT and Harris County Toll Road Authority (HCTRA) without which none of this effort would have been possible.

I would also like to thank all the people who offered me their support and encouragement through the course of this research. A major portion of my two years at Texas A&M were marred by a foot injury and completing this thesis would never have been possible without all the support I got from my friends and peers.

Most importantly, I would like to thank my family (Mummy, Papa and Adithi) who've been there for me every step of the way. Their unwavering support, constant encouragement and unconditional love helped me get through this thesis and complete my master's degree at Texas A&M University.

NOMENCLATURE

AVI	Automated Vehicle Identification
GDOT	Georgia Department of Transportation
GPLs	General Purpose Lanes
GPS	Global Positioning Systems
HOV	High Occupancy Vehicle
HOT	High Occupancy/Toll
HCTRA	Harris County Toll Road Authority
MLs	Managed Lanes
RP	Revealed Preference
RR	Reliability Ratio
SHRP	State Highway Research Program
SH	State Highway
SOVs	Single Occupancy Vehicles
SP	Stated Preference
SR	State Road
TxDOT	Texas Department of Transportation
USDOT	United States Department of Transportation
VOR	Value of Travel Time Reliability
VOT	Value of Travel Time

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Understanding and modeling traveler behavior are the cornerstone of transportation planning. Good planning, in turn, results in sustainable investments in infrastructure and increased economic competitiveness. Travelers make travel decisions based on their understanding and perception of different influencing factors such as the value of time (VOT), comfort, or safety. Predicting how travelers will behave when faced with a choice between a potentially congested but toll-free route and an uncongested but tolled route is particularly challenging. Part of this challenge comes from not understanding the value travelers place on the more reliable travel times offered by the tolled route. A substantial effort is underway by the U.S. Department of Transportation (USDOT) through the Strategic Highway Research Program 2 (SHRP 2) to incorporate travel time reliability into the planning process. In addition, many stated-preference surveys have been undertaken to estimate travelers' value of reliability. Despite these efforts, researchers are still not sure how travelers perceive or value travel time reliability.

Managed lanes (MLs), a component of congestion management, are defined as highway facilities or a set of lanes in which operational strategies are implemented and managed (in real time) in response to changing conditions to preserve unimpeded flow. Often MLs operate alongside free general-purpose lanes (GPLs) in order to allow travelers to choose between the two lanes. High-occupancy vehicles (HOVs) are encouraged in these systems by being allowed to use the tolled lanes either toll free or at a reduced rate.

These types of MLs are becoming increasingly popular in the United States and are now present in multiple cities across the country.

Katy Freeway Dataset

Houston's Katy Freeway is one such ML facility that became operational in 2008. This project tries to understand how travelers value travel time reliability using empirical travel data from Katy Freeway travelers in Houston. The dataset consists of records generated from automated vehicle identification (AVI) sensors placed at regular intervals along the freeway. The data were processed to generate travel times and identify lane choices for every single trip identified on the freeway. This means that not only are the average travel times of the vehicles on the roadway known, but so is the travel time of a particular vehicle on any of its Katy Freeway trips through the identification of its unique transponder. This dataset is unique because Katy Freeway is one of the few freeways that have both tolled and free lanes, and that have AVI readers on both sets of lanes. These data were combined with crash data, lane blockages, weather, and toll rates—the many factors that could potentially impact travel time and travel time reliability. This provided an unmatched dataset of travel conditions on GPLs and MLs.

This dataset was then used to calculate the reliability of travel times on the freeway. Attributes such as time, toll, and reliability were used to run multinomial discrete-choice models and to study the relative importance of these variables in the decision-making process. The travel behavior witnessed provided insight into how much travelers are willing to pay for the travel time savings and reduced travel time variability of the MLs.

Research Objectives

The design of Katy Freeway (two MLs and at least four GPLs in both directions) provides an ideal real-world environment to study how travelers choose between faster, more reliable tolled lanes and congested, less reliable untolled lanes over an extended period of time. This understanding can assist planners in making better decisions to provide sustainable and economically viable transportation options.

The emphasis of this project was understanding and modeling how travelers behave when given a choice between more reliable tolled lanes and less reliable untolled lanes on a daily basis. The project had the following objectives:

1. Calculate the travel behavior of individual travelers on both the toll lanes and GPLs on a freeway using empirical data. Use these data to calculate toll, travel time and travel time reliability for trips on Katy Freeway

2. Gather data on factors that influence how travelers make lane-choice decisions.

3. Estimate an empirical value of travel time reliability using data obtained in objectives 1 and 2.

4. Estimate an empirical value of travel time that is separate from the value of travel time reliability.

Thesis Outline

The organization of the thesis is as follows: Chapter 2 provides a detailed background on the existing research in the area of value of time and value of travel time reliability. It also discusses how this thesis overcomes some of the shortcomings of the existing methodologies. Chapter 3 sheds light on the different data types and their sources used for this research. This is followed by Chapter 4 which discusses the methodology and algorithm that was used to create a trip data set from the raw data. Chapter 5 details the results of the study and inferences derived from these results. Finally, Chapter 6 provides a summary of the thesis by discussing the important conclusions of the study and providing recommendations for future research.

CHAPTER II

LITERATURE REVIEW

Travel time can be defined as the time elapsed when a traveler travels between two (distinct) spatial positions (Carrion and Levinson, 2012). Travel time is typically easily understood because it is a one-dimensional quantity. Travel time reliability, on the other hand, is a concept related to the unpredictability of travel time between two spatial points. It is a measure of the spread of the travel time distribution. In simple terms, the greater the variation in travel time between two points, the less reliable it is and vice versa. Therefore, the concept of travel time (un)reliability is used interchangeably with travel time variability in transportation literature. According to Wong and Sussman (1973), this unpredictability of travel time can be attributed to:

- Variations between seasons and days of the week.
- Variations because of change in travel conditions due to weather, accidents, etc.
- Variations attributed to each traveler's perception.

The VOT and value of reliability (VOR) are measures of travelers' willingness to pay for reducing their travel time and improving the predictability (i.e., reducing travel time variability) of their trip. These are two fundamental factors influencing travel behavior and demand. Therefore, considerable efforts have been made to better understand these factors to improve the transportation planning and decision-making process.

Value of Travel Time

The literature on the value of passenger travel time is extensive and well developed. The earliest studies on the VOT date back to the 1960s (Becker, 1965; Beesley, 1965; Oort, 1969). Values of travel time have most often been determined by estimating mode-choice models and evaluating the marginal rates of substitution between the costs and travel times of the alternative modes. The VOT increases as travelers shift from congested to uncongested travel (Small et al., 1999).

Cherlow (1981) listed various studies conducted on the evaluation of the VOT. The estimated VOT varied from as low as 9 percent of the wage rate to as high as 140 percent of the wage rate. He suggested that there is no single VOT that can be applicable to all people in all circumstances. A study by Lam and Small (2001) estimated the average VOT to be $22.87/hour, or 72 percent of the average wage rate. Patil et al. (2011) estimated the VOT for different situations on MLs, including one normal and six urgent situations. Patil et al. found that travelers place a higher value for travel time savings when in an urgent travel situation than in a normal situation—well over 100 percent of the wage rate.

Few studies in the recent literature try to estimate the VOT on MLs. A study by the Georgia Department of Transportation (GDOT) using a stated-preference survey estimated the VOT of passenger car travelers to be in the range of $7/hour to $15/hour. GDOT also observed that the VOT varied with the type of vehicle. Drivers of six-axle vehicles valued travel time savings at a higher price than drivers of passenger cars

(GDOT, 2010). A more recent study on I-25 travelers in Miami by the Florida Department of Transportation estimated the VOT as 49 percent of the hourly wage, with a range of $2.27/hour to $79.32/hour and a mean value of $32/hour (Perk et al., 2011).

Value of Travel Time Reliability

Research attempting to quantify the VOR is relatively new. Although it has received increased attention, the procedure for quantifying the VOR and the estimated values are still a topic of debate (Carrion et al., 2012). No acceptable valuation exists thus far because existing valuations of the VOR must be examined in light of the underlying assumptions of the study. Three distinct theoretical frameworks have been examined to understand the value of travel time reliability:

- Centrality dispersion.
- Scheduling delays.
- Mean lateness.

The centrality-dispersion framework, in a transportation context, is based on the notion that both expected travel time and travel time variability are sources of disutility:

$$U = \gamma_1 \mu_T + \gamma_2 \sigma_T \quad (1)$$

Where: U = expected utility.

μ_T = expected travel time.

σ_T = dispersion measure of travel time distribution.

$\gamma_{1,2}$ = coefficients.

Therefore, the traveler is minimizing the sum of the two terms: the expected disutility of travel time and the travel time variability of the trip.

The scheduling-delay framework is based on the costs associated with early or late arrival relative to arrival time constraints (e.g., work start time). The travelers' intrinsic choice of a preferred arrival time is the point of reference that delimits if an arrival is early or late. Scheduling decisions made for a given probability density distribution of trip delay and the associated costs are used to determine how travelers value travel time reliability.

The mean-lateness framework is based on the expected utility paradigm. Used primarily for transit, the framework consists of two elements: schedule journey time and mean lateness at arrival. The schedule journey time is the time between scheduled arrival and scheduled departure. The lateness is the sum of lateness at boarding and lateness at arrival. Train fares are added to calculate the marginal rates of substitution between temporal quantities and travel cost.

Among the three, centrality dispersion has been the most popular among researchers. Most of these studies looking at VOR from a centrality dispersion point of view can be categorized into stated-preference studies or revealed-preference studies.

Stated-Preference Studies

To date, stated-preference (SP) techniques have proven to be the most valuable tool for estimating the value of travel time reliability. In their survey, Black and Towriss (1993) asked respondents to choose between distinct options with a varying spread of travel times, mean travel times, and travel costs. They found travel time variability to be a significant factor although the magnitude was found to be smaller than the mean VOT (0.55 times VOT). Black and Towriss also introduced the concept of a reliability ratio (RR), given by:

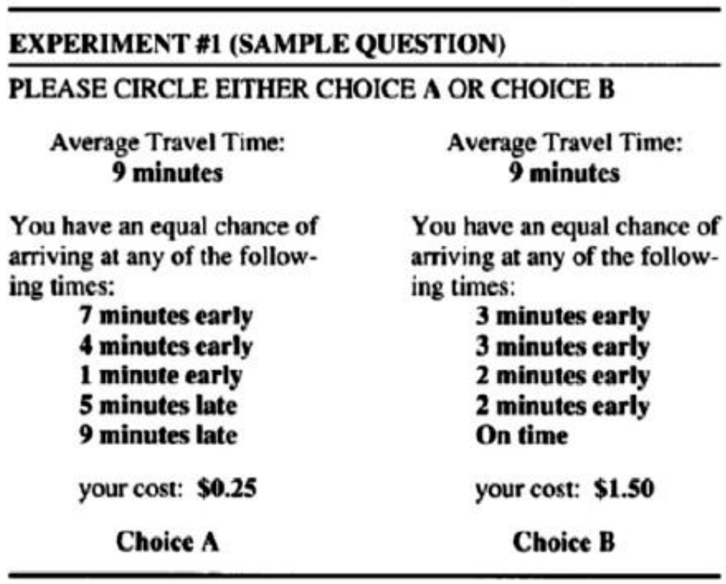
$$RR = \frac{VOR}{VOT} \tag{2}$$

Small et al. (1999) used mean-variance models, scheduling models, and combined models to estimate the VOR. In their survey design, the concept of travel time variability was presented using a text-only format (see Figure 1). The respondents were asked to choose between two scenarios with the same average travel time but different costs and different travel time distributions. Researchers found the VOR on average to be 3.22 times the VOT in congested conditions.

In the late 1990s and 2000s, more attention was devoted to designing better presentations of questions that included variability. Hensher (2001) used bar diagrams to present the concepts of time and reliability to survey respondents. He divided the total travel time into free flow, slowed down, stop/start, and uncertainty allowance (see Figure 2). This was because he was more concerned about the values that travelers assign to each distinct component of travel time rather than general travel time reliability. Three scenarios were provided with varying travel time components and travel costs.

Copley and Murphy (2002), through their qualitative research, found that histogram presentations could present a large volume of information and were understood with little effort by the respondents (see Figure 3). In their survey, the respondents were presented with two choices with varying arrival time distributions.

Tseng et al. (2009) compared the presentation of variability in various studies by conducting face-to-face interviews with subjects to study their understanding of the concepts. Researchers found that the format used by Small et al. (1999) (shown in Figure 1) was understood by most of the respondents and was the most preferred form of presentation.


EXPERIMENT #1 (SAMPLE QUESTION)

PLEASE CIRCLE EITHER CHOICE A OR CHOICE B

Choice A	Choice B
Average Travel Time: **9 minutes**	Average Travel Time: **9 minutes**
You have an equal chance of arriving at any of the following times:	You have an equal chance of arriving at any of the following times:
7 minutes early	**3 minutes early**
4 minutes early	**3 minutes early**
1 minute early	**2 minutes early**
5 minutes late	**2 minutes early**
9 minutes late	**On time**
your cost: **$0.25**	your cost: **$1.50**
Choice A	**Choice B**

Figure 1: Text-Only Presentation of Variability from Small et al. (1999).

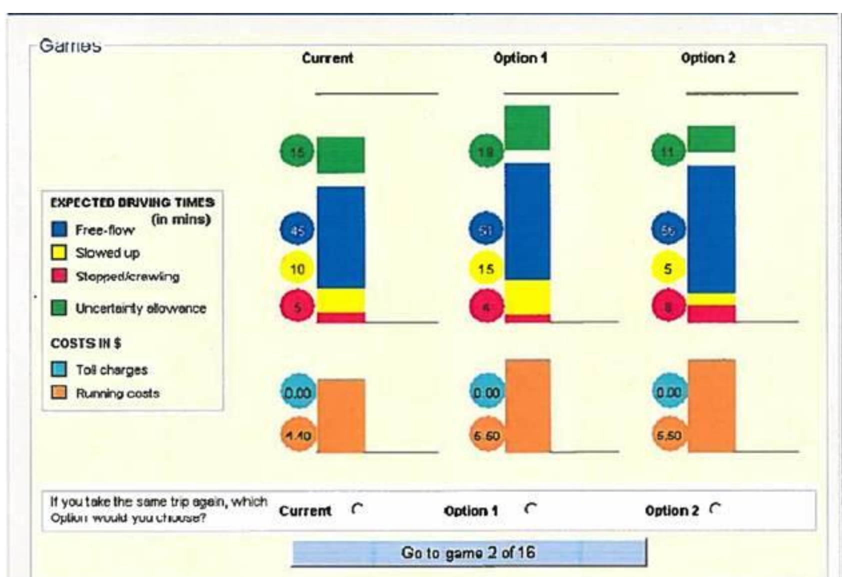


Figure 2: Bar-Diagram Presentation of Variability from Hensher (2001).

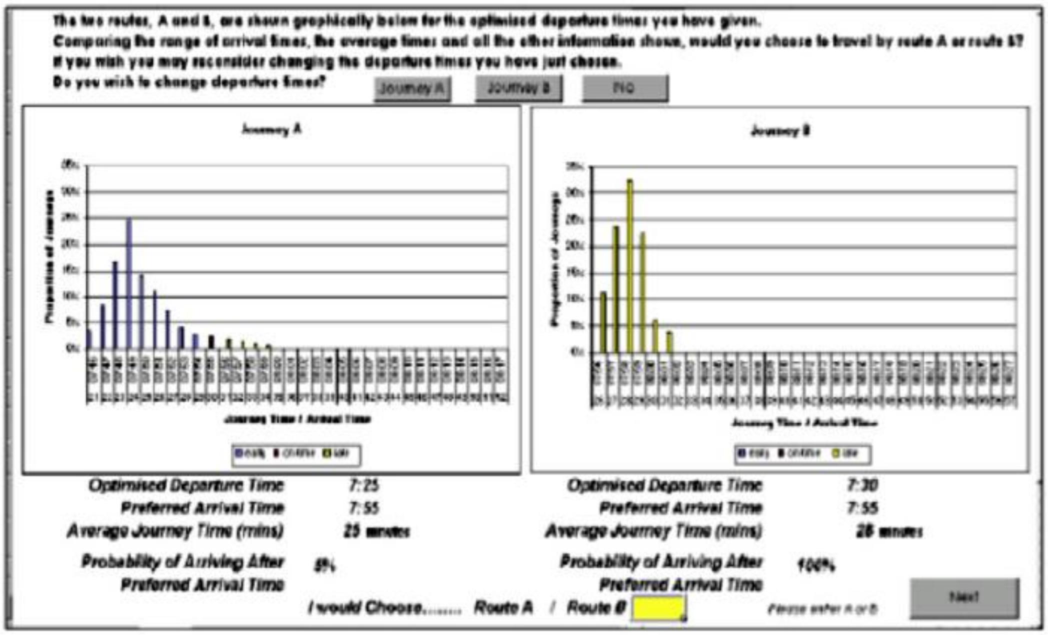


Figure 3: Histogram Presentation of Variability from Copley and Murphy (2002).

Li et al. (2010) derived the value of reliability, scheduling costs, and reliability ratios in the context of Australian toll roads. They used the SP survey to study how travelers in Australia make trade-offs between different levels of travel times and reliability with tolls and vehicle-running costs. They found a mean estimate for early expected schedule delay to be AUS $24.1/hour and a mean estimate for late expected schedule delay of AUS $38.86/hour. The mean VOR from the mean-variance model was AUS $40.39/hour. Unlike in other studies, their study focused on both commuters and non-commuters. A recent study by Tilahun and Levinson (2010) found that the travelers' value travel time reliability was very close to their VOT.

Carrion and Levinson (2012) pointed out that in most SP studies, researchers have not validated whether the respondents' understanding of the abstract situation matches the analysts' intentions of the abstract situation. Therefore, it is difficult to ascertain which measures of travel time variability are more plausible than others. Also, the survey design may affect the reliability estimates due to the difficulty of presenting the concept of travel time reliability to subjects without any statistical background.

Devarasetty et al. (2012) studied the value travelers place on travel time reliability by comparing SP survey data and actual usage data of Katy Freeway travelers. For the survey, approximately half of the respondents received questions in picture format, while the other half received questions in the text-only format. Each question in the survey had four travel choices:

- Drive alone on GPLs.
- Drive alone on toll lanes.
- Carpool on GPLs.
- Carpool on toll lanes (see Figure 4).

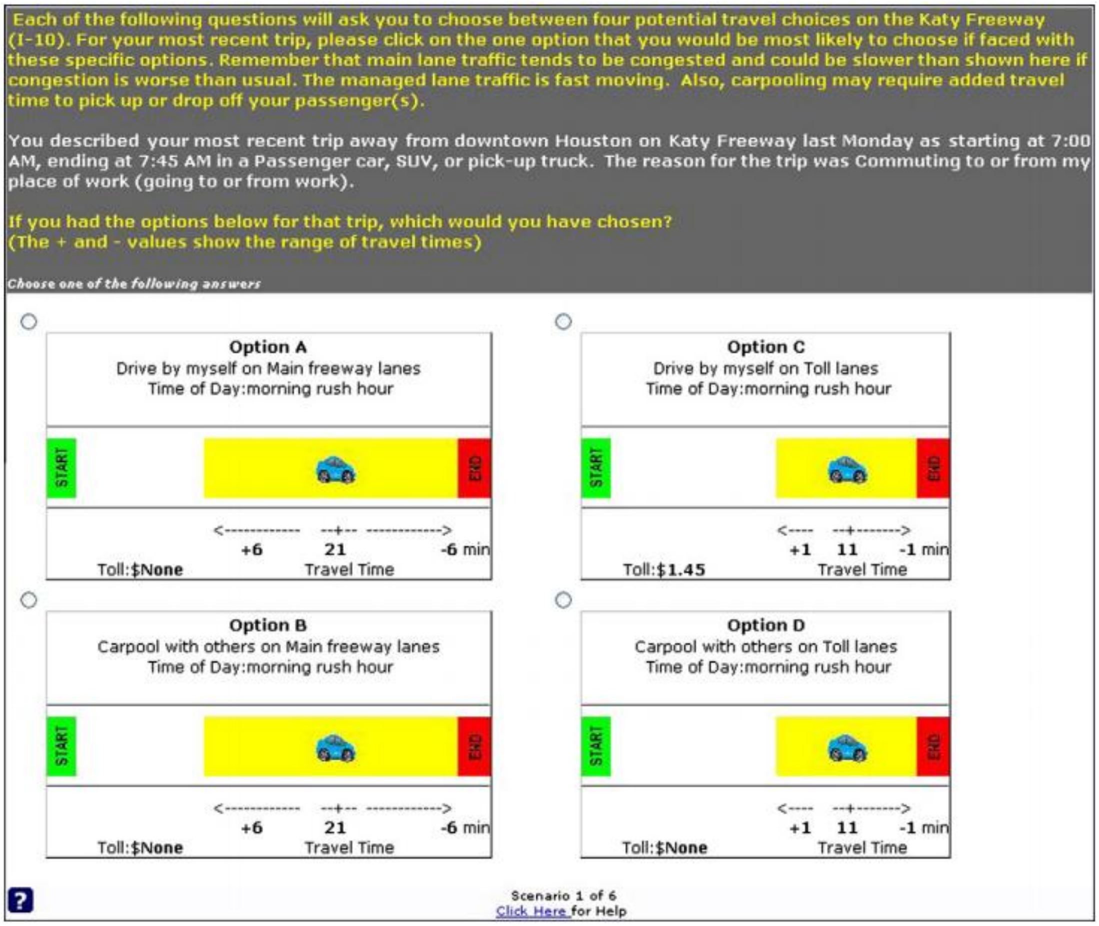


Figure 4: A Typical Scenario in Picture Format from Devarasetty et al. (2012).

Trips on the toll lanes (which had a toll of $1.45 if driving alone and were free if carpooling) were shown to be faster and more reliable than trips on the GPLs. The respondents' choices were similar for both formats, implying that both formats were similarly understood. Three stated-choice experiment design techniques were tested to develop logit models: Bayesian (Db) efficient, random level attribute generation, and adaptive random approach. The Db-efficient design was superior to the other two techniques. Researchers found the combined estimate of the VOT and VOR based on the SP survey (Db-efficient design) to be $50/hour, which was very close to the estimated VOT of $51/hour from actual usage. This implied the calculated VOT using actual usage data includes the value travelers place on reliability plus the value of travel time savings on the MLs.

Concas and Kolpakov (2009) reviewed the literature on the VOT and VOR and recommended that the VOR be estimated at 80 to 100 percent of the VOT under ordinary travel circumstances with no major travel constraints. The most recent investigation of VOR undertaken by the Netherlands Institute for Transport Policy Analysis (2013) used a large SP survey to determine that the VOR ranged from 40 percent of the VOT for commute trips to 110 percent of the VOT for business trips.

Revealed-Preference Studies

Revealed-preference (RP) studies to estimate VOR are few in number in comparison to SP studies. The primary reason is the lack of experimental settings in the real world that provide an opportunity to observe travelers choosing between routes with different travel time reliabilities for a toll. RP studies can be divided into two categories based on how travel time data are measured:

- Subjective travel time measurements, which use travel time as reported by the respondents.

- Objective travel time measurements, which use travel time as obtained from devices such as global positioning systems (GPS) and loop detectors.

A few early studies on the value of reliability using actual traveler usage data were from State Route (SR) 91 in Orange County, California. Travelers chose between a free (less reliable) and variably tolled (more reliable) route. Lam and Small (2001) measured the value of reliability using 1998 loop detector data as well as surveying a sample of the travelers. Using loop detector data, Lam and Small were able to estimate the average travel time within 15-minute intervals. Through the survey of their sample group (533 respondents), Lam and Small were able to gather information about travelers' most recent weekday trip. Unfortunately, the loop detector data were from one year prior to the survey. Therefore, the researchers had to approximately adjust the travel times by using the trends observed in the congestion data. Also, the travel times obtained using the loop detector data were averaged over 15 minutes and were therefore not representative of the specific travel time of a trip maker. Lam and Small found that the best-fitting models represented travel time by the difference between the 90th percentile and the median. In their best models, the value of reliability was \$15.12/hour for men and \$31.91/hour for women. These values were 48 percent and 101 percent, respectively, of the average wage rate in the sample. Because the average wage rates of the sample were used, these figures are not necessarily representative of how much each individual values travel time reliability relative to his or her wage rate.

Liu et al. (2011) attempted to analyze the travel behavior and preferences of SOV users of HOT lanes using field data collected by point traffic sensors and transponder toll tags. The researchers found that the difference in VOT distribution for frequent and infrequent SOV users was not statistically significant meaning frequent and infrequent users value travel time savings similarly. Furthermore, logit-like models were developed to study lane choice behavior for pre-congestion, congestion and post-congestion phases. For the models, travel time reliability was measured using a probability based approach where reliability was the probability of travel time being less than a prespecified travel time for a given time of the day. The travel time coefficient had negative coefficients during all the three phases indicating an increase in utility accrued with a decrease in travel time. Based on the models, it was concluded that SOVs care more about travel time savings and travel time reliability for the pre and post-congestion phases owing to the relatively lower toll rates. However, during the congestion phase tolling functions as a signal to attract more traffic into the HOT lane so as to avoid downstream congestion. That is, a higher toll rate indicates to users that the downstream traffic in the GP lane is fairly congested and thus encourages HOT lane use.

Small et al. (2005, 2006) combined both RP (actual preferences of lane choice) and SP (hypothetical scenarios to better understand lane choice) observations of travelers on SR 91. The data were collected using telephone interviews and mail-back surveys. The researchers noted that RP data collected using surveys are affected by perception errors and therefore can never be completely representative of the real-world scenario. Also, 522 individuals took the RP survey, and 633 took the SP survey, but only 55 respondents took both. Therefore, combining data from different individual may have introduced errors in the study. The researchers calculated the median VOR using the RP data of travelers in Los Angeles and estimated the median VOR to be 85 percent of the average wage rate ($19.56/hour). The researchers also found the heterogeneity in travel time and reliability to be significant. Brownstone and Small (2005), using the data from SR 91 and I-15 high-occupancy toll (HOT) lanes, estimated the VOR to be 95 to 140 percent of the median travel time.

In another study, Carrion and Levinson (2013) estimated the value placed on the HOT lanes because of improvements in travel time reliability from a GPS-based experimental design. the researchers recruited 18 regular commuters on the I-394 MnPass lanes in Minnesota and equipped them with GPS devices. For the first two weeks, these commuters were instructed to travel on each of the three alternatives (HOT lanes, untolled lanes [GPLs], and nearby signalized arterials) and then were given the opportunity to travel on their preferred route after experiencing each alternative. The researchers found that reliability was valued highly in each of the models but was valued differently according to how it was defined (standard deviation, shortened right range, and interquartile range). Though the study used RP data, the low number of participants in the experiment design may have biased the results.

Variation in Studies

From the discussion in the previous sections, it is clear that large variations in estimates have been found across different studies. These differences in the valuation of travel time reliability are a key problem in comparing estimates across studies. Tseng and Verhoef (2008) classified the main differences into the following categories:

- Data type (RP, SP, and joint RP and SP).
- Scheduling versus reliability measures.
- Various travel time reliability measures (e.g., standard deviation and interquartile range).
- Travel time unit.
- Presence of heterogeneity (observed and unobserved).
- Choice dimensions (e.g., mode and route).

The data type differences are primarily the difference in SP, RP, and actual usage studies. SP studies are affected by the perception of the VOR of respondents and their understanding of abstract concepts. Often in RP and actual usage studies, there is simply not enough variation in tolls to be able to empirically determine the influence of cost on decision making (Devarasetty et al., 2012). Scheduling and reliability measures are closely related and thus obscure the contribution of each other in model estimates. Differences in reliability measures lead to variation in results. Different reliability measures, such as standard deviation, difference in 90th and 50th percentiles, etc., have been used in studies. Heterogeneity in the subjects of studies (e.g., socio-economics attributes) can lead to varying results across studies. These factors interact with the travel time, reliability, and cost terms, making it difficult to estimate the valuation ratios. Finally, differences in the choice behavior of travelers between mode, route, and departure time may influence the estimation of the VOT and VOR. This means the lack of homogeneity in the experiment design of various studies potentially leads to variations in the results.

About This Research

The methodology in this thesis aims to address most of the issues faced in previous studies of travel time reliability. Using actual freeway usage data, instead of surveys, eliminates the concerns of survey-based studies. It also provides the actual choices travelers made, thus eliminating perception errors induced in SP survey studies. Also, the data allow for an estimate of the value of travel time that is separate from the value of travel time reliability.

As discussed, some RP studies have tried to incorporate actual usage data to validate or improve their model estimates. The main problem with these studies is the collection of usable travel time data. Loop detectors, in-field measurements (e.g., driving in similar time periods), and GPS devices have been used. Loop detectors require several assumptions and processing to estimate travel time. In-field measurements are easier to get but do not reflect actual travel times. GPS devices measure very detailed commute-level data but are difficult to obtain. This research overcomes these problems by using highly accurate AVI data to identify trip times on both MLs and GPLs. Moreover, previous studies that used RP data and actual usage data had to approximate travel times using algorithms and could not obtain the actual travel time of the specific respondent being surveyed in the RP survey (Lam and Small, 2001; Small et al., 2005, 2006). In this project, accurate travel times for each individual traveler with a transponder identification (ID) number were available (on both the MLs and GPLs), and no approximation of travel times was needed.

Another reason for variation in results among different studies is the influence of time of day over travel time. Measures from off-peak hours may differ from those during the peak hours. The dataset for this project was used to identify travel times during all hours of the day with no exception. This enabled researchers to generate separate datasets based on the time of day and develop different models to understand variations in results, if any, due to the time of day.

CHAPTER III

DATA

Katy Freeway

Katy Freeway is one of Houston's major highways and connects the city of Katy in the west to downtown Houston in the east. The highway has a total length of 40 miles and was constructed in the 1960s. The initial design had three lanes per direction and two frontage lanes to accommodate approximately 80,000 vehicles per day.

In the late 1990s and early 2000s, traffic volumes started reaching three times the volume the highway could serve, which resulted in chronic congestion levels lasting up to 11 hours a day. This led to the Texas Department of Transportation (TxDOT) undertaking a major five-year reconstruction project for a 12-mile section of the freeway between just west of SH 6 and the I-10/I-610 interchange. The project broke ground in 2003 and was completed in October 2008. This $2.79 billion project was partially funded through a combination of state funds, federal funds, and toll receipts.

The reconstruction widened the 12-mile stretch to provide up to six GPLs in each direction and two variably priced MLs in each direction in the median of the highway. Figure 5 is a detailed map of Katy Freeway. The two lanes in the middle running in both directions are MLs with four entry and four exit points in both directions. HOVs with two or more occupants and motorcycles do not have to pay a toll during the HOV-free hours but have to pay the same toll rate as single-occupancy vehicles (SOVs) during all other hours. HOV hours are Monday through Friday 5 a.m. to 11 a.m. and 2 p.m. to 8 p.m. This is done to encourage ride sharing and increase vehicle occupancy. HOVs must also be in the HOV lane (inside lane) of the MLs to avoid the toll during the HOV-free hours. GPLs have no tolls at any time. In general, the MLs provide faster and more reliable travel times compared to the GPLs.

The freeway has three tolling plazas (near the cross streets of Eldridge, Wilcrest, and Wirt) for toll collection from vehicles. All toll collection at these booths is done electronically, and vehicles need to have transponders in order to use the toll lanes. All vehicles passing through any toll booth are identified with the help of the transponder and are charged a toll depending on the time of the day and the toll plaza. The operation and toll collection for the freeway are handled by the Harris County Toll Road Authority (HCTRA).

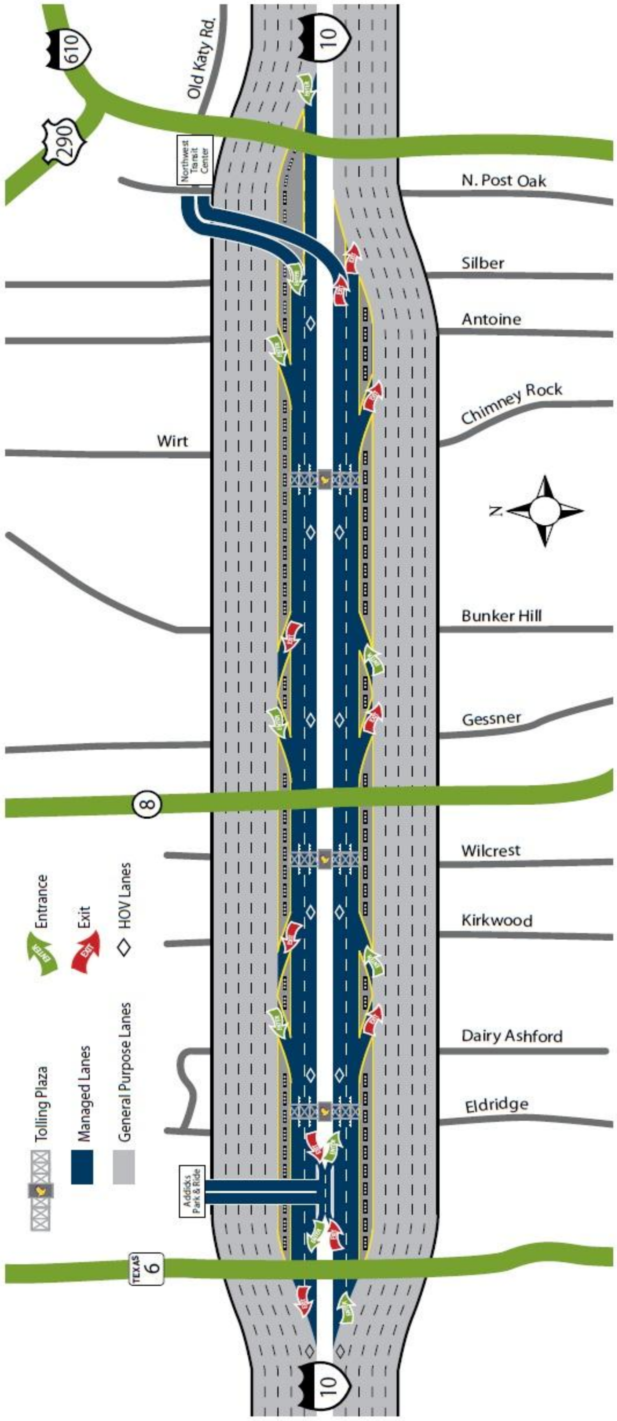


Figure 5: Katy Freeway.
(Source: Harris County Toll Road Authority)

To legally use the MLs, SOVs must pay a toll, which varies by time of day. The toll rates that were in effect during the period of analysis are shown in Table 1.

Table 1: Toll Schedule on Katy Managed Lanes (April 2012).

Time Period	Toll Plaza		
	At Wilcrest	At Wirt	At Eldridge
Peak Period (7–9 a.m. Eastbound and 4–6 p.m. Westbound)	$1.20	$1.20	$1.60
Shoulder (6–7 a.m. and 9–10 a.m. Eastbound and 3–4 a.m. and 6–7 p.m. Westbound)	$0.60	$0.60	$0.80
Off-Peak Period (All Other Times)	$0.30	$0.60	$0.40

All along the freeway, in both directions, AVI sensors are placed along the MLs, GPLs, and frontage road lanes. The section of highway examined in this research had 38 readers owned and operated by TxDOT (see Figure 6). Each sensor is assigned a unique number, which is used to identify the location and direction of travel of vehicles passing the sensor. Only vehicles with a valid transponder ID are detected at these sensors. Upon detecting a vehicle, the sensor records the time of detection and the unique transponder ID of the vehicle. For this project, these data and the location of the sensors were used to identify vehicle trips on the freeway. Each transponder ID was assigned a unique random number, and the original ID was deleted. In this way, it was impossible to trace the records back to the driver who made the trip. Figure 6 shows the location of the AVI sensors.

Data Sources

AVI Data

The AVI or automatic vehicle identification data contain all sensor detection records of vehicles with transponders on Katy Freeway for 2012. These data were obtained from TxDOT. Each record has the time stamp, the sensor number, and the unique vehicle transponder ID. The data were processed and used to identify trips on the MLs and GPLs. The dataset was modified to make it impossible for the researchers to identify a specific traveler. The dataset was combined with the HCTRA toll data, and then each transponder ID was assigned a unique random ID number. When the researchers were satisfied the randomizing process worked correctly, the original data and the randomizing code were deleted, making it impossible to identify the original transponder ID and the person who took a particular trip.

For 2012, 225,118,768 records were obtained from the 38 readers on Katy Freeway. This amounted to 1,993,347 unique transponder ID detections for the entire year. For April, 870,819 unique transponder IDs were detected on the freeway, and 4,496,918 trips were identified in both directions.

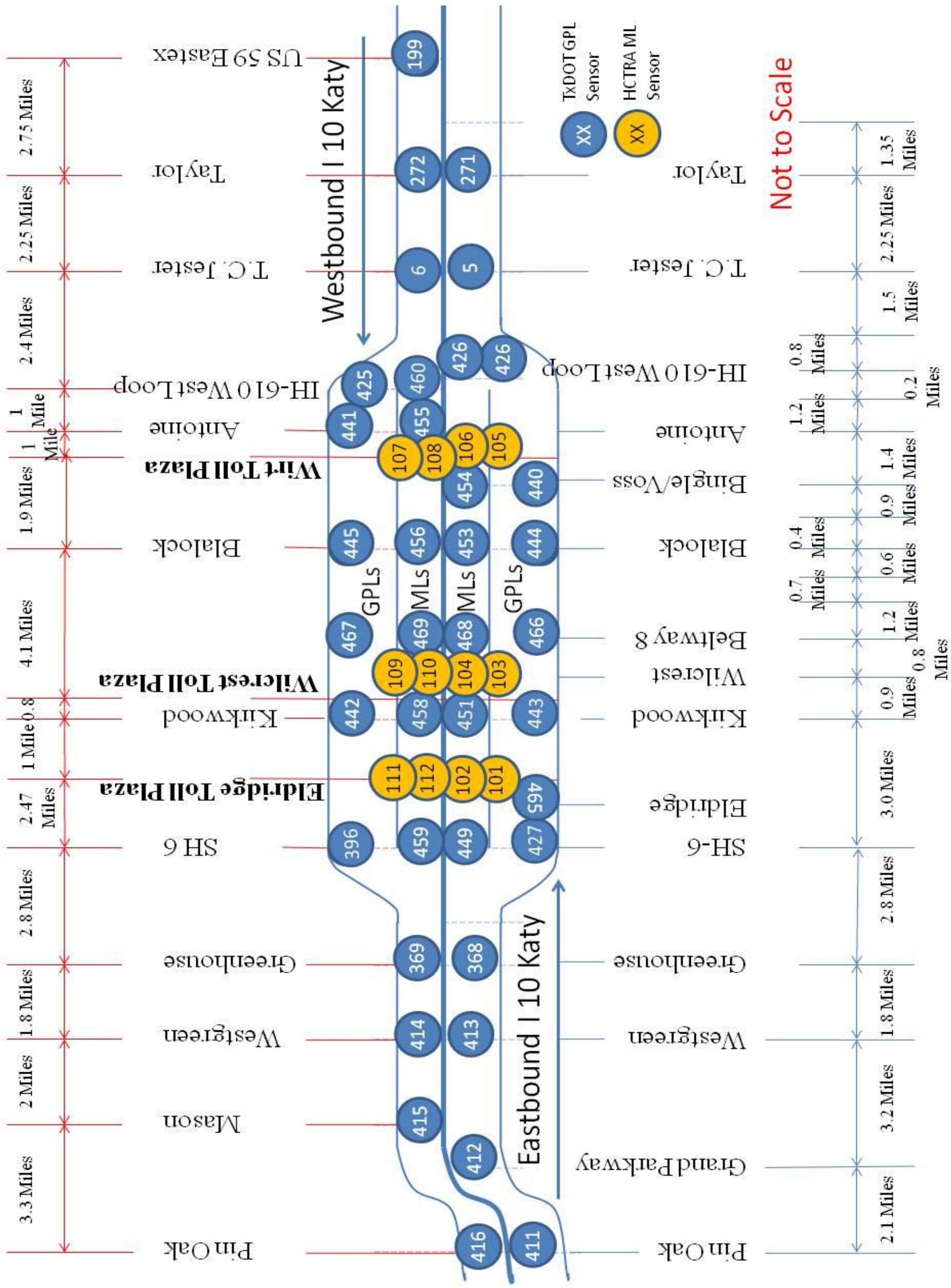


Figure 6: Katy Freeway AVI Sensor Locations.

HCTRA Toll Data

The tolling dataset for 2012 was obtained from HCTRA, which used the data to charge the vehicle the appropriate toll rate based on the toll schedule. The dataset contained all vehicles with valid transponder IDs that were detected at the toll plazas along the MLs on Katy Freeway. Each record in the dataset contains the time stamp of detection, the unique transponder ID of the vehicle detected, location, toll plaza ID number, and lane ID number. For the purpose of this project, these data were used to supplement the AVI detections in order to better identify trips along the MLs. Based on the location, toll plaza ID and lane ID number unique sensors numbers were assigned to each of the toll sensors. All the twelve toll sensors on the Katy freeway were assigned sensors numbers from 101 to 112. These toll sensor locations are shown in Figure 6. In the trip identification algorithm, these data were also used to assign the correct toll rate to each trip identified on the MLs or properly identify non-tolled vehicles in the case of toll-free HOVs on the MLs. For 2012, 14,769,730 toll detection records were generated, with 1,310,043 during April 2012.

Crash and Lane Closure Data

TxDOT provided a dataset containing information pertaining to all incidents (including crashes) and resulting lane closures on Katy Freeway for 2012. This dataset was included to factor in the impact of lane closures on travel time, travel time reliability, and decision making of the traveler. In theory, the impact of crashes on travel time and travel time reliability was already accounted for with the data collected on those two variables. However, radio and Internet announcements of lane closures may have impacted travel, and so lane closures were included as independent variables.

The dataset contained information about the type of incident, type of lanes affected (ML, GPL, or frontage), number of lanes affected, duration for which lanes were affected, and location of the incident. Researchers then identified the nearest sensors to the site of the incident. For 2012, 1,178 incidents were recorded. Of these, 36 records had incomplete information and were therefore deleted. All the incidents led to at least one lane being blocked on the freeway system. For April, 121 incidents were recorded, and these were the only ones included in the analysis.

Weather Data

Data pertaining to hourly rain information were obtained from the National Climatic Data Center. The dataset contained the hourly rainfall level in inches near Katy Freeway. A variable to identify heavy rain was incorporated in the dataset when recorded rainfall was greater than 0.4 inches in an hour. In April 2012, there were four such hours with more than 0.4 inches of precipitation.

Research Methodology

For this thesis, only one month of data was used due to the volume of trips and the computational power needed to deal with the analysis. April was chosen because it was a fairly typical traffic month, it was before the summer when traffic decreases on Katy Freeway, and it was well before the toll changed in September 2012.

Achieving the objectives of this research required a set of empirical data that included a majority of the attributes used by travelers to choose between a priced and reliable set of lanes (MLs) versus an unpriced (or lower priced) but less reliable set of lanes (GPLs). Using the datasets mentioned in the previous section, a new dataset for April containing all the identified trips and all the possible attributes that could be ascertained was created. Some of the trip attributes included were the random ID of the vehicle, travel time, toll paid, travel time standard deviation, time of day when the trip was made (peak or off-peak), and trip length. For all the trips identified on the freeway, a second trip was generated on the lanes not chosen (i.e., GPL for ML trips and ML for GPL trips). The attributes for this dummy trip were calculated using the information available from trips that were made on those same lanes at the same time (a 10-minute window).

Since it is unclear how travelers perceive travel time reliability, different variables can be used to represent travel time reliability. For this thesis, three different measures of travel time reliability were used which are as follows:

- the standard deviation of travel time
- coefficient of variation (standard deviation of travel time/ mean travel time)
- relative standard deviation (standard deviation of travel time/ actual travel time)

In other words, the higher these measures, the less reliable the trip would be.

Obtaining a dataset that includes every attribute a traveler might use in choosing between these two set of lanes would be impossible. Thus, a methodology that allows for some unknown attributes and their impacts was needed. Logit models were used to estimate lane-choice behavior (GPL versus ML). A logit model is a type of regression analysis used for predicting the outcome of a categorical dependent variable based on one or more predictor variables. In this thesis the dependent variable is binary (ML or GPL), and therefore the logit model is called a binary logit model. The following is an example of the type of logit model used in this research:

$$U_{GPL} = \beta_{TT} TravelTime_{GPL} + \beta_{TTR} TravelTimeReliability_{GPL} + \beta_{GLB} GPLsBlocked$$

$$\begin{aligned} U_{ML} = \beta_{ML} + \beta_{TollML} Toll_{ML} + \beta_{TT} TravelTime_{ML} + \beta_{TTR} TravelTimeReliability_{ML} + \\ \beta_{LB} MLsBlocked + \beta_{rain} Rain \end{aligned} \quad (3)$$

Where: U_i = utility derived by choosing lane i.

β_i = coefficient associated with attribute i.

GPL = general-purpose lane.

ML = managed lane.

Toll = toll paid on the toll lane.

TT = travel time.

TTR = travel time reliability.

GLB = GPLs blocked.

MLB = MLs blocked.

The VOT and value of travel time reliability could then be obtained by comparing the relative importance of attributes related to travel time, travel time reliability, and toll in determining lane choice.

CHAPTER IV

DATA ANALYSIS

This chapter briefly discusses the algorithm that was used to identify the trips and trip attributes. It also outlines all the assumptions that were made and their impact on the analysis, and clearly explains the final dataset that was obtained. The SAS program used for the thesis is attached in Appendix A.

Cleaning, Merging, and Randomization of Data

The first step of the analysis was to clean the raw data. No records with incomplete information were found in the AVI and HCTRA data, but a small number of duplicate entries were found and removed.

As discussed in Chapter 3 toll sensors were given sensor numbers to enable merging with the AVI sensor data. In the original dataset, each toll booth was identified by a plaza ID. In order to enable the merge, each toll booth was assigned a sensor number instead of the plaza ID to more clearly resemble the AVI sensor data. All attributes other than the time stamp, sensor number, and transponder ID were removed.

After the merge, all transponder IDs were assigned a unique random ID, and the original transponder ID was deleted. This ensured no trip could be mapped back to the original transponder ID. This randomized dataset was used for all subsequent analysis.

Records corresponding to random IDs that were detected only once (that is, a single location) were deleted because no trip could be identified by a single detection. Therefore, a dataset with only random IDs that were detected more than once was created. After these initial steps, the total number of records (individual transponder reads) for the whole year was 225,118,768. For April, the total number of records was 19,383,952.

Trip Identification

Records were sorted in chronological order and according to random ID. Therefore, consecutive detections, one after the other, for the same random ID were chained together to trace a trip through the freeway. For example, a specific vehicle transponder identified at readers 427, then 465, then 443, and finally 466 within a given time period was converted into a single trip entering the freeway at reader 427 and exiting at reader 466 (refer to Figure 6). If the time difference between two consecutive detections for the same random ID was greater than 10 minutes, the two detections were considered to be part of two different trips.

Trip times were calculated by taking the difference in time between the first detection and the last detection. Similarly, trip length was calculated by measuring the distance between the location of the first and last sensors. Distances between all sensors were calculated using Google Earth and are shown in Figure 6.

A toll was assigned to a trip that had at least one detection at one of the toll plazas. The tolls were assigned according to the time of detection and the corresponding toll value in the toll schedule. No toll was applied to the trip if the vehicle was detected on the HOV lane portion of the MLs during the HOV-free hours. The total toll for the trip was equal to the sum of tolls paid along the trip at up to three different toll booths.

For this thesis, only sensors within the 12 mile stretch of the freeway with managed lanes were used for identifying trips. This means, sensors such as sensor 369, 412, 199, 5 etc. were deleted from the dataset. This was done because, outside of the managed lanes section, toll lane and GPL users share the same roadway and therefore face the same traffic conditions within a 10 min interval. Moreover, the paid users pay to save travel time and get better reliability only on the section of the freeway where the toll lanes are separate from the general purpose lanes.

Estimation of Trip Reliability Measures

As mentioned in the previous chapter, three different measures of travel time reliability were used for this research. Computing the standard deviation of travel was the first step to compute each of these three measures. The travel time standard deviation was calculated per 15-minute interval for each sensor pair using all the trip times during that 15-minute interval for that sensor pair. Despite the large number of trips in the dataset, there were some 15-minute periods when there were too few trips to determine the standard deviation on one set of lanes (GPLs or MLs). When there were very low volumes of traffic (two or fewer vehicles per 15-minute period) on the lanes, the standard deviation for that sensor pair was allowed to be zero. When there was traffic on the lanes but that particular sensor pair did not experience enough trips, the standard deviation was estimated by using the regression equation shown in Equation 4:

$$\sigma_{x,y} = 0.48 \times \sum(\sigma_i) + 2.20 \times S + 6.37 \qquad (4)$$

Where: $\sigma_{X,Y}$ = standard deviation of travel speeds between sensor pair X, Y of the roadway.

σ_i = standard deviation of speeds for all adjacent sensor pairs located between sensor pair X,Y.

S = number of adjacent sensor pairs between sensor pair X,Y.

For example, the standard deviation for a trip that starts at sensor 413 and ends at sensors 444 while going through sensors 368 and 443 is given by Equation 5:

$$\sigma_{413,444} = 0.48 \times (\sigma_{413,368} + \sigma_{368,443} + \sigma_{443,444}) + 2.20 \times 3 + 6.37 \qquad (5)$$

This regression equation was calculated by using data from all the sensor pairs, for all those 10-minute periods where at least five trips were identified between the given sensor pair. For example, in the first step, if five trips were observed during a 10 minute period between sensor pair 413 and 444 (refer to equation 5) then standard deviation was computed for this pair using the travel times of the five identified trips. In the second step, standard deviation for the intermediate sensor pairs i.e. 413-368, 368-443 and 443-444 was also computed using the trips identified between these sensors pairs. The sum of the standard deviation of these individual intermediate sensors pairs calculated in step two and the actual standard deviation calculated in step one were then used to develop a regression equation that related the actual observed standard deviation to the sum of the individual standard deviations of the intermediate sensor pairs.

The entire month's data were used to develop the regression equation. A total of 2,576,194 trip sensor pairs had five or more trips identified during a particular 10 minute interval. Data from all these trip sensor pairs were used to develop the regression equation. The regression model had an R-squared value of 0.49. A number of different models were tried to improve the R-squared value, but no significant improvement was observed. Different models for the peak, off-peak, and shoulder were developed. Step-wise regression was tried with a large number of trip variables, but in both cases, the complexity of the model increased significantly without a commensurate increase in the R-squared value. A model with only the sum of the standard deviation of speeds was also tried, but the model had a lower R-square than the one for Equation 5.

This estimated standard deviation was used to calculate the other measures of reliability used for this thesis. The coefficient of variation is given by:

$$C_v = \frac{\sigma}{\mu} \tag{6}$$

Where: C_v = Coefficient of variation.

σ = Standard deviation of travel time.

μ = Mean travel time.

The other measure of reliability used for this thesis was standard deviation of travel time relative to the actual trip time experienced. This measure was used because travelers may tolerate a relatively higher standard deviation for a longer trip compared to a shorter trip. This means that a standard deviation of 1 minute for a 10 minute trip would be favorable than a trip with a standard deviation of 1 minute and travel time of 2 minutes. Therefore, the standard deviation relative to travel time was given by:

$$\sigma_{rel} = \frac{\sigma}{T_t} \tag{7}$$

Where: σ_{rel} = Standard deviation relative to actual travel time

σ = Standard deviation of travel time.

T_t = Actual travel time experience for the trip.

Alternate Trip Generation

An alternate trip was generated to develop attributes for the lane that was not chosen. This means for every trip made on the MLs, an alternate trip was created on the GPLs and vice versa. This was necessary to create attributes for the alternate choice that could then be used in the logit model.

The alternate trip was created such that it passed through the same section of the freeway but on the other set of lanes (ML for GPL trips and GPL for ML trips). The start time of the trip was assumed to be the exact same as the start time of the original trip. The length of the alternate trip could vary a small amount (up to 0.3 miles) depending on the relative location of sensors on both sets of lanes. For alternate trips created on the toll lanes, tolls were assigned depending on the number of toll booths present in the section of the freeway on which the alternate trip was generated.

The travel time for the alternate trip was determined by taking the average of travel times between the given sensor pair during the 15-minute interval in which the actual trip was made. When no trips were observed between a sensor pair in a given interval, then average travel times were used. These average travel times were calculated using actual trips that occurred on these lanes during the same time frame (off-peak, shoulder and peak) on days with travel time data. Table 2 summarizes average speeds used to determine travel times. Standard deviations were calculated the same way.

Table 2: Speed Comparison by Period.

Period	**Average Speed on the Toll Lanes (in mph)**	**Average Speed on the GPLs (in mph)**
Peak Period (7–9 a.m. Eastbound and 4–6 p.m. Westbound)	53.1	40.8
Shoulder (6–7 a.m. and 9–10 a.m. Eastbound and 3–4 a.m. and 6–7 p.m. Westbound)	61.8	55.1
Off-Peak Period (All Other Times)	70.9	65.3

Note: The speed comparisons are for the entire trip identified, which may include short parts of the trip that are outside the 12 miles of the toll lane.

Lane Changes

Trips that switched between the GPL and ML, or vice versa, during the ML segment of Katy Freeway could not be used. This was because it was impossible to determine the exact location of the lane switch because vehicles were only detected at the sensors, which had fixed locations. This meant travel time savings were impossible to correctly estimate. However, this does not include the longer ML trips that were detected on GPL sensors outside of the 12-mile ML section. For example, ML trips that were detected at GPL sensors 411, 415, 6, 271, etc. (refer to Figure 6) were not deleted. Due to the deletions, the total number of trips in the dataset decreased by from 3,560,636 to 3,076,838.

Trips on the HOV Lane Part of the MLs

The objective of the research was to compare how travelers choose between tolled and toll-free lanes. All trips made on the HOV lanes during the HOV-free times do not pay any tolls but enjoy the same benefits as tolled trips on the toll lanes. Analyzing this third option was beyond the scope of this project and so all trips on the HOV lanes during HOV-free hours were removed. This reduced the total trips for April from 3,076,838 trips to 2,781,355 trips.

Additional Attributes

In addition to the travel time, time of day, day of week, toll, and travel time variability, additional attributes were included to better explain the trip parameters. A variable for heavy rain on the freeway was included (1 if rain was greater than 0.4 inches in that hour and 0 otherwise).

Lane blockages can have an adverse impact on travel time and reliability. Traffic moving from the blocked lanes to the other lanes has the potential to disturb travel time on all the lanes. Therefore, variables were created to account for any lane blockages observed during the trip. The type (ML, GPL, or frontage) and number of lanes blocked were included in the dataset. The number of lanes blocked was attributed to a trip only if the trip was identified to go through the impacted area on the freeway.

Trips were also classified according to the time at which the trip was made. Trips were classified as peak-hour trips, peak-shoulder trips, and off-peak trips. The peak hours were 7 to 9 a.m. eastbound and 4 to 6 p.m. westbound on weekdays. The peak-shoulder hours were 6 to 7 a.m. and 9 to 10 a.m. eastbound and 3 to 4 p.m. and 6 to 7 p.m. westbound on weekdays. All the other times were classified as off-peak.

Final Dataset

The final dataset had two records per trip identified. The two records represented the two potential choices for the trip: one that was made and one on the lanes not chosen. The trip parameters included in the final dataset were the random ID, lane choice, trip time, trip time standard deviation, total toll paid, trip length, lane blockages and heavy rain. These attributes were the independent variables used in the logit models.

The final dataset was used to create four datasets based on the time when the trip was made. The first dataset had all the trips including weekends that were identified during the entire month. The second data set had only the peak period trips (7–9 a.m. Eastbound and 4–6 p.m. Westbound). The third data set consisted of only the shoulder period trips (6–7 a.m. and 9–10 a.m. Eastbound and 3–4 a.m. and 6–7 p.m. Westbound). Both the peak and shoulder period data sets contained only the weekday trips. No weekend trips were included in this dataset. The fourth data set was created for off-peak period trips and this data set consisted of all trips not made during the peak and shoulder periods including the weekend trips. Models were developed separately for each of these four datasets.

CHAPTER V

RESULTS AND INFERENCES

After identifying the trips and the trip attributes, the data was analyzed to understand travel behavior on the MLs. First, statistical measures relating to trip attributes were developed to gain preliminary insight into the observed travel trends. Then, logit models were estimated using the available data to further analyze the travel behavior. This chapter discusses the results obtained.

Trips were categorized into peak-period trips, shoulder trips, and off-peak-period trips (see Table 3).

Table 3: Classification of Trips by Time of Day.

Time Period	Paid Trips*		GPL Trips		Total Trips**
	Count	Percentage	Count	Percentage	
Peak Period (7–9 a.m. Eastbound and 4–6 p.m. Westbound)	71,194	2.56	259,915	9.34	331,019
Shoulder (6–7 a.m. and 9–10 a.m. Eastbound and 3–4 p.m. and 6–7 p.m. Westbound)	34,845	1.25	272,103	9.78	306,948
Off-Peak Period (All Other Times)	58,027	2.09	2,085,271	74.97	2,143,298
Total Trips	164,066	5.90	2,617,289	94.10	2,781,355

* Paid trips on the MLs made by SOVs and HOVs during non-HOV-free hours (for all weekdays of April 2012)

** Total trips excludes trips made by vehicles without transponder IDs, trips on the HOV lanes during HOV-free hours, and trips detected on both MLs and GPLs in the ML portion of the freeway.

The percentage of toll-paying trips made on the MLs decreases from the peak periods to the shoulders to the off-peak periods. This can be attributed to the decreasing travel time savings and decreasing difference in travel time reliability, as shown in Table 4. The numbers for travel time savings and travel time reliability improvement in Table 4 are for the entire trip identified regardless of the length of the trip.

Table 4: Average Travel Time Savings and Travel Time Reliability Savings by Period.

Period	Average Travel Time Savings on the Toll Lanes (in Minutes)	Average Travel Time Reliability Improvement on the Toll Lanes (in Minutes)
Peak Period (7–9 a.m. Eastbound and 4–6 p.m. Westbound)	3.49	0.25
Shoulder (6–7 a.m. and 9–10 a.m. Eastbound and 3–4 a.m. and 6–7 p.m. Westbound)	1.87	0.14
Off-Peak Period (All Other Times)	0.81	0.15

Average travel time savings for toll-paying trips on the MLs decreases from 3.49 minutes during the peak periods to 0.81 minutes during the off-peak periods. Similarly, the difference in travel time reliability (standard deviation) between MLs and GPLs falls from 0.25 minutes during the peak periods to 0.15 minutes during the off-peak periods though a lowest reliability improvement of 0.14 minutes is observed during the shoulder hours. Considering only the toll-paying trips on the MLs (HOV trips were not included), it would be logical that travelers would be less willing to pay and use the MLs during the off-peak and shoulder periods as compared to peak periods. Also, based on the speed comparison by period (Refer to Table 2) it can be summarized that average speeds increase from the peak periods to shoulders to off-peak periods for both MLs and GPLs. However, the difference in speeds between MLs and GPLs decreases significantly from the peak to shoulder to off-peak periods. This could have further dissuaded travelers from using the toll lanes during off-peak periods even though the toll rates were lower during the off-peak periods.

Figure 7 categorizes the identified trips based on the length of the trips. It must be noted that these trip lengths are for the trips identified on the freeway using sensor detections. The actual trip length could be longer and may have extended beyond the 12 mile stretch of the Katy freeway.

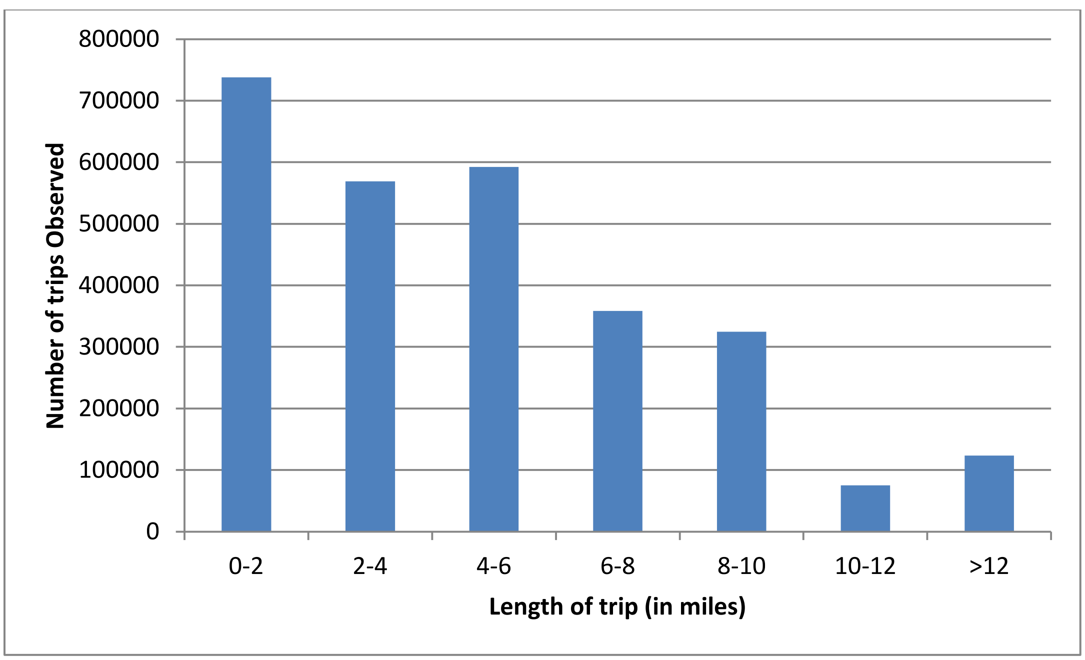


Figure 7: Length of Trips Identified

Based on the travel behavior observed on the freeway, route-choice models were developed, as shown in Equation 6. The multinomial discrete-choice modeling procedure in Statistical Analysis System (SAS) was used to generate the models. These are standard multinomial logit models that attempt to identify what role several factors play in the traveler's lane choice (GPL or ML). It must be noted that due to the large data set and the complexity of the models being developed random variation of taste across travelers was not modeled i.e. coefficients for model attributes were assumed to be the same across all travelers. Therefore, independent variables such as travel time did not have a random component. Also, personal information (e.g., gender, age, or income) was not available, and only information on travel conditions, as noted in the previous sections, was used in the models. Equation 8 is one of the models that were developed for the given data. Simpler models with only time and toll or travel time reliability and toll and more complex models including lane closure and precipitation attributes were also developed.

$$U_{GPL} = \beta_{TT} TravelTime_{GPL} + \beta_{TTR} TravelTimeReliability_{GPL}$$

$$U_{ML} = \beta_{ML} + \beta_{TollML} Toll_{ML} + \beta_{TT} TravelTime_{ML} + \beta_{TTR} TravelTimeReliability_{ML} \quad (8)$$

Where: GPL = general-purpose lane.

ML = managed lane.

Toll = toll paid in the ML.

TT = travel time.

TTR = travel time reliability.

As mentioned in Chapter4, models with different sets of variables were developed for the whole month (including weekends), peak periods only, shoulders only, and off-peak periods only. Table 5 shows the results of several models that were generated to potentially explain the travel behavior observed on the freeway.

Intuitively, an increase in toll and travel time should lead to a decrease in utility. In models with only time and toll, negative coefficients were observed for both time and toll. Based on the value of the coefficients, the marginal rate of substitution of time with toll increased from off-peak periods ($1.53/hour) to shoulder periods ($6.05/hour) to peak periods ($9.05/hour). For the entire month, this marginal rate of substitution was $5.45/hour. This indicates that the relative importance of time with respect to cost (toll) is higher for travelers during the peak periods as compared to the off-peak periods. The addition of an alternate specific coefficient (ASC) to the same time and toll model changes the coefficients drastically. Large negative values of the ML ASC are observed in all models, and the coefficient of toll has a positive value. A positive coefficient for toll in the model means that higher the toll, more likely travelers are to choose the toll lanes over the general purpose lanes. This seems counterintuitive as an increase in toll (cost) should lead to decrease in utility derived. A possible explanation for this could be that users choose the toll lanes more during the peak and shoulder periods (as noted in Table 3) when the congestion is high. However, these periods have higher toll rates compared to all other times of the day. The combined effect of these two factors is captured in the model, which associates higher toll rates with greater probability of the tolls lanes being chosen.

Table 5: Logit Models Based on Empirical Data.

Variable	Coefficient (Standard Error)			
	All Month	**Peak Period**	**Shoulder**	**Off-Peak**
Model: $U_{TOLL} = B_1 \times time + B_2 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times toll$†				
Time	−0.23 (0.001)	−0.08 (0.001)	−0.12 (0.002)	−0.17 (0.003)
Toll	−2.53 (0.004)	−0.53 (0.003)	−1.19 (0.005)	−6.64 (0.010)
Model: $U_{TOLL} = ASC\_ML + B_1 \times time + B_2 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times toll$				
ASC_ML	−4.30 (0.005)	−3.80 (0.013)	−3.56 (0.013)	−5.28 (0.010)
Time	−0.17 (0.001)	−0.19 (0.001)	−0.09 (0.002)	−0.09 (0.003)
Toll	1.05 (0.003)	0.82 (0.005)	0.86 (0.008)	2.45 (0.013)
Model: $U_{TOLL} = B_1 \times std + B_2 \times toll$, $U_{GPL} = B_1 \times std + B_2 \times toll$				
Std Dev. (Unreliability)	2.22 (0.008)	1.62 (0.012)	0.56 (0.015)	1.33 (0.013)
Toll	−1.71 (0.003)	−0.22 (0.002)	−0.94 (0.004)	−5.99 (0.008)
Model: $U_{TOLL} = ASC\_ML + B_1 \times std + B_2 \times toll$, $U_{GPL} = B_1 \times std + B_2 \times toll$				
ASC_ML	−4.19 (0.005)	−3.16 (0.012)	−3.63 (0.012)	−5.24 (0.010)
Std. Dev. (Unreliability)	0.31 (0.008)	0.55 (0.013)	0.02 (0.018)	0.28 (0.014)
Toll	1.32 (0.003)	0.95 (0.005)	1.06 (0.007)	2.61 (0.012)
Model: $U_{TOLL} = B_1 \times time + B_2 \times std + B_3 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times std + B_3 \times toll$				
Time	−0.36 (0.001)	−0.13 (0.001)	−0.18 (0.002)	−0.31 (0.003)
Std. Dev. (Unreliability)	2.73 (0.008)	1.96 (0.013)	1.02 (0.017)	1.62 (0.014)
Toll	−2.36 (0.003)	−0.43 (0.003)	−1.16 (0.005)	−6.45 (0.010)
Model: $U_{TOLL} = ASC\_ML + B_1 \times time + B_2 \times std + B_3 \times toll$, $U_{TOLL} = B_1 \times time + B_2 \times std + B_3 \times toll$				
ASC_ML	−4.23 (0.005)	−3.62 (0.013)	−3.55 (0.012)	−5.23 (0.010)
Time	−0.19 (0.001)	−0.20 (0.001)	−0.10 (0.002)	−0.12 (0.003)

Table 5 Continued

Variable	Coefficient (Standard Error)			
	All Month	Peak Period	Shoulder	Off-Peak
Std. Dev. (Unreliability)	0.64 (0.009)	0.90 (0.014)	0.25 (0.019)	0.43 (0.014)
Toll	1.05 (0.003)	0.81 (0.005)	0.86 (0.008)	2.41 (0.013)

† The toll for GPL trips was 0.

Similar to time and toll, an increase in the measure of (un)reliability (standard deviation) was also expected to lead to a decrease in the utility. But positive coefficients for standard deviation were observed for all models that included standard deviation of time along with the toll. The addition of an ASC to the standard deviation and toll model also led to high ASC values. This again may be because of the inability of the variables used to completely explain the observed behavior. The models with time, toll, and standard deviation gave similar results with negative coefficients for time and toll when developed without the ASC, and gave high ASC values when developed with the ASC.

As has been noted in transportation literature, toll variables generally have negative coefficients in choice models. As the toll goes up, the utility derived goes down. In the models with the ASC, the tolls have positive coefficients in most cases. The toll rate schedule is the same throughout the month except the peak hours and shoulders (four hours) during the weekdays. During the peak hours and shoulders, the toll rates are higher but constant from day to day. Therefore, this lack of variability in the toll schedule may have led to the model not being able to correctly gauge the impact of toll on lane choice. Also, during the off-peak hours when the tolls are lower, the percentage of trips on the MLs goes down drastically. The lower toll rates during the off-peak hours are not able to attract the same percentage of trips on the toll lanes as observed during the peak hours. This means a smaller percentage of the total trips data available during the off-peak hours is from the toll lanes. Therefore, this relative scarcity of toll lane data during the off-peak periods could have hampered the results.

The behavior of travelers on the MLs and how they perceive travel time reliability are also not clear. Even though the standard deviation of travel time is a good measure of the spread of travel time, results show that it may not be how travelers perceive travel time reliability. Therefore, two other measures of reliability, namely the coefficient of variation and the standard deviation of travel time relative to actual travel time were tested in the models. Table 6 summarizes the results of these models.

Table 6: Logit Models with Different Measures of Reliability.

Variable	Coefficient (Standard Error)			
	All Month	Peak Period	Shoulder	Off-Peak
Model: $U_{TOLL} = B_1 \times C_v + B_2 \times Toll$, $U_{GPL} = B_1 \times C_v + B_2 \times Toll$				
Coeff. Of Variation(C_v)	4.19 (0.022)	5.69 (0.064)	0.04 (0.018)	3.50 (0.037)
Toll	-2.03 (0.003)	−0.35 (0.002)	-0.99 (0.004)	-6.25 (0.008)
Model: $U_{TOLL} = B_1 \times time + B_2 \times C_v + B_3 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times C_v + B_3 \times toll$				
Time	−0.22 (0.001)	-0.06 (0.001)	-0.12 (0.002)	-0.16 (0.003)
Coeff. Of Variation(C_v)	4.08 (0.021)	4.64 (0.066)	0.02 (0.019)	3.38 (0.038)
Toll	-2.48 (0.004)	-0.47 (0.003)	-1.19 (0.005)	-6.53 (0.009)
Model: $U_{TOLL} = B_1 \times \sigma_{rel} + B_2 \times Toll$, $U_{GPL} = B_1 \times \sigma_{rel} + B_2 \times Toll$				
Rel. Std. Dev. (Unreliability)	4.24 (0.021)	5.28 (0.061)	-0.04 (0.063)	3.57 (0.037)
Toll	-2.02 (0.003)	-0.35 (0.002)	-0.99 (0.004)	-6.24 (0.008)
Model: $U_{TOLL} = B_1 \times time + B_2 \times \sigma_{rel} + B_3 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times \sigma_{rel} + B_3 \times toll$				
Time	−0.21 (0.001)	−0.06 (0.001)	-0.12 (0.002)	−0.14 (0.003)
Rel. Std. Dev. (Unreliability)	4.00 (0.021)	4.27 (0.062)	-0.39 (0.066)	3.40 (0.038)
Toll	-2.46 (0.004)	−0.47 (0.003)	-1.19 (0.005)	-6.49 (0.010)

Coefficients for both the measures of reliability tried were found to be positive in all the models developed. These counter-intuitive results for different measures travel time reliability measures indicate that the valuation travelers' place on travel time reliability and how they perceive travel time reliability are far more complicated than what has been reported in the existing literature. As discussed in the literature review section, most studies have suggested that traveler's value travel time reliability. An increase in reliability (decrease in unreliability) leads to an increase in the utility derived which in turn has an effect on their decision-making. However, the empirical data of Katy freeway users analyzed in this study does not give the same results as most of the survey based SP and RP studies. Therefore, travel behavior of users of the Katy freeway indicates that users may not value travel time reliability as explained by the utility theory, where an increase in travel time reliability of a lane leads to an increase in the utility derived from choosing that lane which in turn increases the probability of that lane being chosen.

It could also be that most travelers may perceive travel time reliability different from the measures tried in this research. For instance for a traveler, one bad experience (high delay) on the MLs after having paid a toll may be enough to cause him or her to not use the MLs again for a long time. These others measures of reliability may be able to better capture how travelers' value travel time reliability. Also, at the time of making the choice, the decision maker does not have complete information about the travel time and the travel time reliability. Therefore, a large part of the decision-making process is governed by historical data (travel times experienced in the past), which is not accounted for in the model.

It is also possible that Katy Freeway travelers do not make lane choices based on the variables used in this model such as travel time savings or travel time reliability. This may seem like irrational behavior on the part of the travelers, but other factors, such as familiarity in daily travel (lack of willingness to change lanes regularly) and a desire to avoid weaving in and out of MLs, may be more important to these travelers.

Due to the limited entry and exit points in and out of the MLs, travelers may have preferred to stay away from the MLs on short trips and avoid the hassles of weaving in and out of the MLs. Therefore, lane choice could have varied based on the length of the trip. In order to study whether the length of the trip had any impact on how travelers make choices and perceive travel time reliability, the dataset was divided into three categories based on the length of the trips (short, medium, and long). Trips less than 3.5 miles were categorized as short trips, trips between 3.5 and 8 miles as medium trips, and trips greater than 8 miles as long trips. Separate models were developed based on the three datasets, but little difference was found based on trip length. Tables 7 and 8 summarize the results of the models generated.

Table 7: Lane Choice Based on Trip Length

Trip Type	Paid Trips		GPL Trips		Total Trips
	Count	Percentage	Count	Percentage	
Short (<3.5 Miles)	31,757	1.14	1,309,208	47.07	1,290,775
Medium (3.5 to 8 Miles)	90,201	3.24	1,104,752	39.72	1,416,620
Long (>8 Miles)	42,108	1.52	203,329	7.31	424,900
Total Trips	164,066	5.90	2,617,289	94.10	2,781,355

As shown in Table 8, developing models based on the length of the trip had little impact on the models. Based on these results, it is clear that travel time, toll, and travel time reliability are not sufficient to explain lane-choice behavior on Katy Freeway. Models were also generated by including additional attributes such as lane blockages and rain, but the results were basically the same.

Table 8: Logit Models Based on Trip Lengths.

Variable	Coefficient (Standard Error)		
	Short	Medium	Long
Model: $U_{TOLL}= B_1 \times time + B_2 \times toll$, $U_{GPL}= B_1 \times time + B_2 \times toll$			
Time	-0.35 (0.003)	−0.01 (0.001)	−0.12 (0.002)
Toll	−5.39 (0.010)	−1.54 (0.004)	−0.78 (0.006)
Model: $U_{TOLL}= ASC\_ML + B_1 \times time + B_2 \times toll$, $U_{GPL}= B_1 \times time + B_2 \times toll$			
ASC_ML	−4.91 (0.010)	−4.04 (0.007)	−3.30 (0.011)
Time	−0.29 (0.003)	−0.10 (0.001)	−0.12 (0.002)
Toll	0.82 (0.007)	1.22 (0.004)	0.91 (0.007)
Model: $U_{TOLL}= B_1 \times std + B_2 \times toll$, $U_{GPL}= B_1 \times std + B_2 \times toll$			
Std. Dev. (Unreliability)	0.98 (0.016)	3.53 (0.010)	2.72 (0.015)
Toll	−4.77 (0.008)	−0.87 (0.003)	−0.13 (0.004)
Model: $U_{TOLL}= ASC\_ML + B_1 \times std + B_2 \times toll$, $U_{GPL}= B_1 \times std + B_2 \times toll$			
ASC_ML	−4.80 (0.010)	−3.87 (0.007)	−3.02 (0.011)
Std. Dev. (Unreliability)	0.15 (0.021)	0.55 (0.012)	1.33 (0.017)
Toll	1.12 (0.006)	1.39 (0.004)	1.23 (0.006)
Model: $U_{TOLL}= B_1 \times time + B_2 \times std + B_3 \times toll$, $U_{GPL}= B_1 \times time + B_2 \times std + B_3 \times toll$			
Time	−0.48 (0.004)	−0.08 (0.002)	−0.21 (0.002)
Std. Dev. (Unreliability)	1.55 (0.016)	3.57 (0.011)	3.04 (0.017)
Toll	−5.42 (0.010)	−1.03 (0.005)	−0.62 (0.006)
Model: $U_{TOLL}= ASC\_ML + B_1 \times time + B_2 \times std + B_3 \times toll$, $U_{TOLL}= B_1 \times time + B_2 \times std + B_3 \times toll$			
ASC_ML	−4.87 (0.010)	−3.91 (0.007)	−3.03 (0.012)
Time	−0.31 (0.003)	−0.11 (0.001)	−0.16 (0.002)

Table 8 Continued

Variable	Coefficient (Standard Error)		
	Short	Medium	Long
Std. Dev. (Unreliability)	0.62 (0.020)	0.71 (0.012)	1.57 (0.018)
Toll	0.80 (0.007)	1.22 (0.004)	0.89 (0.008)

Additional models with interaction terms and squared terms were then developed to see if the observed behavior could be explained better. Two separate combinations of the time and toll attributes, namely toll multiplied by time and toll divided by time were used as interaction terms in the model with travel time, standard deviation of travel time and toll attributes. Squared terms for each of time, toll and reliability were used to study if a second degree polynomial regression could better explain the observed behavior. The developed models are summarized in Table 9 and Table 10.

Table 9: Logit Models with Interaction Terms.

Variable	Coefficient (Standard Error)			
	All Month	**Peak Period**	**Shoulder**	**Off-Peak**
Model: $U_{TOLL} = B_1 \times time + B_2 \times toll + B_3 \times (toll \times time)$, $U_{GPL} = B_1 \times time + B_2 \times toll + B_3 \times (toll \times time)$				
Time	−0.13 (0.001)	0.01 (0.001)	−0.02 (0.002)	−0.17 (0.003)
Std. Dev. (Unreliability)	2.31 (0.009)	1.53 (0.014)	0.66 (0.017)	1.77 (0.013)
Toll	−5.55 (0.007)	−1.32 (0.006)	−2.92 (0.011)	-12.67 (0.022)
(Toll × Time)	0.60 (0.001)	0.16 (0.001)	0.32 (0.002)	1.29 (0.003)
Model: $U_{TOLL} = B_1 \times time + B_2 \times toll + B_3 \times (toll / time)$, $U_{GPL} = B_1 \times time + B_2 \times toll + B_3 \times (toll / time)$				
Time	−0.04 (0.001)	−0.01 (0.001)	−0.03 (0.002)	−0.07 (0.004)
Std. Dev. (Unreliability)	2.62 (0.008)	1.55 (0.013)	0.91 (0.017)	2.02 (0.013)
Toll	1.52 (0.005)	0.67 (0.005)	-0.62 (0.009)	0.70 (0.017)
(Toll/Time)	-15.23 (0.026)	-4.60 (0.022)	-7.28 (0.038)	-25.07 (0.078)

Table 10: Logit Models with Squared Terms.

Variable	Coefficient (Standard Error)			
	All Month	Peak Period	Shoulder	Off-Peak
Model: $U_{TOLL} = B_1 \times time + B_2 \times std + B_3 \times toll + B_4 \times toll^2$, $U_{GPL} = B_1 \times time + B_2 \times std + B_3 \times toll + B_4 \times toll^2$				
Time	−0.34 (0.001)	−0.20 (0.001)	-0.03 (0.003)	−0.28 (0.004)
Std. Dev. (Unreliability)	1.68 (0.009)	1.32 (0.015)	0.53 (0.018)	1.04 (0.014)
Toll	-5.83 (0.007)	-2.52 (0.009)	-3.18 (0.011)	-10.51 (0.016)
$Toll^2$	1.69 (0.003)	0.67 (0.002)	0.96 (0.004)	5.88 (0.016)
Model: $U_{TOLL} = B_1 \times time + B_2 \times std + B_3 \times toll + B_4 \times time^2$, $U_{GPL} = B_1 \times time + B_2 \times std + B_3 \times toll + B_4 \times time^2$				
Time	0.61 (0.002)	0.25 (0.003)	0.51 (0.005)	0.68 (0.006)
Std. Dev. (Unreliability)	2.16 (0.008)	1.48 (0.013)	0.51 (0.017)	2.11 (0.014)
Toll	-2.21 (0.004)	-0.36 (0.003)	-1.06 (0.005)	-6.16 (0.009)
$Time^2$	-0.04 (0.001)	-0.01 (0.001)	-0.03 (0.000)	-0.07 (0.001)
Model: $U_{TOLL} = B_1 \times time + B_2 \times std + B_3 \times toll + B_4 \times time^2$, $U_{GPL} = B_1 \times time + B_2 \times std + B_3 \times toll + B_4 \times time^2$				
Time	-0.34 (0.001)	-0.13 (0.001)	-0.18 (0.002)	-0.28 (0.003)
Std. Dev. (Unreliability)	4.70 (0.012)	3.03 (0.022)	3.00 (0.035)	3.18 (0.022)
Toll	-2.27 (0.004)	-0.41 (0.003)	-1.13 (0.005)	-6.24 (0.010)
Std^2	-1.40 (0.006)	0.58 (0.009)	-1.13 (0.018)	-0.99 (0.010)

The use of interaction terms in the time, standard deviation and toll model did not lead to any improvement in the model. As before, negative coefficients were observed for the toll and time attribute while positive coefficient was observed for the standard deviation attribute. Moreover, without any improvement in the model, the presence of the interaction terms makes it harder to explain the models. Similarly, the use of squared terms for toll, time and reliability made little difference to the time, toll and reliability models discussed earlier.

As discussed earlier in this chapter, models with both the travel time and travel time (un)reliability were developed in this thesis. Therefore, it was important to study the correlation between the two attributes. The correlation coefficients for travel time and travel time (un)reliability are shown in Table 11.

Table 11: Correlation Coefficients for Travel Time and Travel Time (Un)Reliability.

All	Peak Period	Shoulder	Off-Peak Period
0.58	0.53	0.65	0.55

A weak positive correlation was found between the two attributes meaning that trips with higher travel times were more likely to have higher travel time variability. This correlation between travel time and travel time reliability could have potentially impacted the models that included both the attributes. However, models with only time and toll without standard deviation of travel time had negative coefficients for time while models with only standard deviation of travel time and toll without time were found to have positive coefficients for standard deviation of travel time. This seems to indicate that travel time and travel time (un)reliability do not impact lane choice behavior in the same way which reconfirms the need to include both attributes separately in lane choice models.

The correlation between travel time and travel time (un)reliability could be one of the reasons why it is difficult to separately analyze the value travelers' place on travel time reliability. For the Katy Freeway, travelers' may look at travel time and travel time reliability synonymously, which in turns makes it more difficult to understand how they perceive travel time reliability separate from travel time.

As the results obtained were counter-intuitive, it was important to ensure the veracity of the results and the accuracy of the algorithm being used to identify trips and trip attributes. Individual trips were randomly picked and traced back to the original randomized dataset to ensure that the trips were being pieced together correctly. Individual trip attributes such as trip time, trip length, and trip time standard deviation were examined for any anomalies or unexpected trends. The toll values applied on the trips were also crosschecked manually with the prevalent toll schedule at the time the trip was made to ensure accuracy. On the modeling side, individual weeks of April were modeled separately to ensure the same results were being obtained for all weeks. A model without any observations in which the trip time standard deviation was approximated using the regression model was also developed. The standard deviation was approximated in cases when there wasn't enough traffic volume (two or less than two trips identified in a 10 min interval between a given sensor pair) detected on the freeway to compute the true standard deviation. Typically, this was observed during the off-peak period and shoulder periods when the percentages of toll trips were low. Therefore, a higher percentage of trips during the off-peak and shoulder periods were eliminated as compared to the peak periods creating a biased sample of the total trip population identified. The developed models are summarized in Table 12. As can be seen there is little difference from the models that were developed by including trips where travel time standard deviation was approximated. The sign of the attribute coefficients were the same as observed in the earlier models whereas the difference in magnitude of these coefficients was because of the biased sample used to develop these models. Therefore, it can be concluded that approximating standard deviations for trips with low traffic volumes had little impact on the model results.

Table 12: Logit Models with Only Trips with True Standard Deviations

Variable	Coefficient (Standard Error)			
	All Month	**Peak Period**	**Shoulder**	**Off-Peak**
Model: $U_{TOLL} = B_1 \times time + B_2 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times toll$†				
Time	−0.18 (0.001)	−0.16 (0.001)	−0.23 (0.003)	−0.14 (0.004)
Toll	−0.58 (0.003)	−0.42 (0.003)	−0.74 (0.005)	−1.64 (0.011)
Model: $U_{TOLL} = B_1 \times std + B_2 \times toll$, $U_{GPL} = B_1 \times std + B_2 \times toll$				
Std Dev. (Unreliability)	0.47 (0.009)	0.35 (0.015)	-0.14 (0.018)	0.33 (0.013)
Toll	−0.25 (0.003)	−0.14 (0.002)	−0.39 (0.004)	−1.32 (0.008)
Model: $U_{TOLL} = B_1 \times time + B_2 \times std + B_3 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times std + B_3 \times toll$				
Time	−0.19 (0.001)	−0.17 (0.001)	−0.24 (0.002)	−0.22 (0.004)
Std. Dev. (Unreliability)	0.80 (0.010)	0.68 (0.013)	0.29 (0.017)	0.65 (0.016)
Toll	−0.54 (0.003)	−0.39 (0.003)	−0.73 (0.005)	−1.61 (0.011)

A large number of trips were used to develop the models in this thesis. One of the shortcomings of developing models with large data is that it is difficult to gauge the significance of a model attribute as most attributes come out to be significant. Also, sometimes the finer nuances of the population are harder to identify in the models. Therefore, models were developed on smaller random sample of the identified trip population to ensure that the attributes were significant and their coefficients were comparable with models developed with the entire data set. Randomly selected two percent of the total population was used to develop these models. In all the models tested the attributes of time, toll and travel time standard deviation were significant with a confidence of 99% and the attribute coefficients were similar to the coefficients obtained for models developed with the entire dataset. Models developed for one such small sample are summarized in Table 13.

Table 13: Logit Models developed for Small Sample Size (2% of Total Trips)

Variable	Coefficient (Standard Error)			
	All Month	**Peak Period**	**Shoulder**	**Off-Peak**
Model: $U_{TOLL} = B_1 \times time + B_2 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times toll$†				
Time	−0.24 (0.023)	−0.09 (0.021)	−0.09 (0.046)	−0.11 (0.059)
Toll	−2.75 (0.075)	−0.63 (0.061)	−1.17 (0.110)	−6.37 (0.059)
Model: $U_{TOLL} = B_1 \times std + B_2 \times toll$, $U_{GPL} = B_1 \times std + B_2 \times toll$				
Std Dev. (Unreliability)	2.12 (0.155)	1.92 (0.025)	0.65 (0.302)	1.41 (0.248)
Toll	−1.90 (0.056)	−0.24 (0.045)	−0.96 (0.085)	−5.82 (0.008)
Model: $U_{TOLL} = B_1 \times time + B_2 \times std + B_3 \times toll$, $U_{GPL} = B_1 \times time + B_2 \times std + B_3 \times toll$				
Time	−0.38 (0.024)	−0.14 (0.024)	−0.14 (0.048)	−0.28 (0.066)
Std. Dev. (Unreliability)	2.66 (0.164)	2.21 (0.274)	0.95 (0.337)	1.70 (0.257)
Toll	−2.58 (0.077)	−0.48 (0.065)	−1.146 (0.110)	−6.21 (0.187)

Lastly, a handful of ML and GPL trips were handpicked and modeled manually. The results were verified with a SAS-generated model to ensure the correct models were being developed. As discussed previously, other personal attributes such as age, gender, income, and attitude toward risk were not available and hence could not be accounted for in the study. The presence of these variables and their interactions with existing variables could have helped better explain the behavior of travelers and their choices.

CHAPTER VI:

CONCLUSIONS AND FUTURE WORK

The objective of this thesis was to understand travel behavior on the Katy freeway using empirical data and calculate a value of travel time reliability separate from the value of travel time using data gathered on lane choices made by the travelers. Lane-choice logit models were developed to study the behavior and attributes that influenced lane-choice decisions.

Based on the data, it was found that the percentage of total trips on the toll lanes was higher during the peak hours. The percentage of toll lane trips decreased during the shoulder periods and further decreased during the off-peak periods. This was in agreement with the decreased travel time saving and decreased travel time reliability saving that existed during the off-peak period.

Models were developed to explain the observed behavior on the freeway. These included various combinations of travelers' travel time, travel time reliability measures, and toll paid. None of the models were able to conclusively explain the observed behavior on the freeway. Most models had positive coefficients for the reliability measures tested, suggesting an increase in travel time variability of an option increased the utility of that option. This is contrary to how rational travelers would perceive travel time reliability. Only in the time and toll model (without an ASC) were negative coefficients observed for both time and toll. These models yielded reasonable values of time of $1.53/hour,

$6.05/hour, and $9.05/hour for off-peak period, shoulder, and peak-period travelers, respectively. Also, it was observed that the addition of an ASC to models led to relatively high ASC values (in magnitude) compared to other attribute coefficients. This implies that there was a weak relationship between lane choice and other trip attributes.

Three separate measures of reliability, (1) standard deviation of travel time, (2) coefficient of variation of travel time and (3) travel time standard deviation relative to the total trip time were tested. All the models developed with these attributes yielded positive coefficients for these reliability measures. That is, the higher the unreliability of a lane the greater the probability of the lane being chosen. These counter-intuitive results suggest that at least in the case of the Katy freeway travelers, the valuation travelers' place on travel time reliability and their perception of travel time reliability are far more complex that what has been reported in existing literature.

Different models were also developed based on the length of the trip to evaluate if behavior changed based on trip length. Little difference was found in the different models; only minor variations in the magnitude of attribute coefficients were found. Models with time-toll interaction terms and travel time squared, toll squared and reliability squared terms were also developed to see if a second degree polynomial equation could help better capture the observed behavior on the freeway in the models. Addition of these terms did not provide any additional insight into the travel behavior or improvements in the model. Models were also developed by adding rain and lane blockage attributes but the results were basically the same.

Overall, it was concluded that travel time, the three measures of travel time reliability and toll were not sufficient to explain the observed lane choice behavior of the Katy freeway travelers. This research based on empirical data does not concur with the existing travel time reliability research, largely based on surveys, which predominantly states that travelers' place a value on travel time reliability while making travel decisions (lane choice, route choice or mode choice). Based on the observed behavior, travelers, at least on the Katy freeway, did not give importance to or value reliability for making lane choice decisions.

The inability of the models to provide more intuitive results could have several potential causes. These can be broadly categorized as data specific and model specific. The lack of sufficient variation in the toll schedule and the low percentage of total trips observed on the MLs during shoulder periods and off-peak periods are issues related to the empirical data used. Also, travel time and travel time standard deviation had to be approximated in cases where there was lack of traffic volume between certain sensor pairs during different times of the day. These factors are related to the empirical data used and could have potentially impacted the results of the models.

A few limitations of the modeling efforts in this research could have also impacted the results obtained. Lack of personal information of travelers meant that the models were developed solely using trip attributes. The inclusion of traveler attributes could have potentially better explained the observed behavior and also enabled the study of the interaction between trip and traveler attributes. As stated earlier, the measures of reliability used may not be the best measure of travelers' perception of travel time reliability. More clarity on how travelers' perceive travel time reliability could have helped develop better models. Lastly, due to computational limitations of running logit models on SAS, random taste variation across travelers could not be incorporated into the model.

Research trying to understand travel time reliability and how traveler's value travel time reliability is in its early stages. First and foremost, further research is needed to understand the measure of travel time reliability that best represents travelers' perception of reliability. The three measures tested in this research may not be the best representation of how travelers' perceive travel time reliability. This could be a potential reason for the observed results. Other measures of reliability that could be considered include covariance of travel time, interquartile ranges of travel time distributions (e.g., the difference between 90th and 50th percentile), number of bad trips in the last 20 trips (e.g., a bad trip could be defined as a trip with a travel time greater than twice the average travel time), and others. A better understanding of the travelers' perception would in turn help better quantify the value, if any, travelers' place on reliability.

In order to study some of these reliability measures it is important to track trips over a longer period of time. Due to computational limitations, only a month of data could be used for this project. But tracking and analyzing trips over a longer period of time would potentially give better insight into how travelers make decisions on the freeway.

REFERENCES

Becker, G.S. (1965). A Theory of the Allocation of Time. *The Economic Journal,* Vol. 75, pp. 493–517.

Beesley, M.E. (1965). The Value of Time Spent in Travelling: Some New Evidence. *Economica*, Vol. 32, pp. 174–185.

Black, I.G., and J.G. Towriss (1993). *Demand Effects of Travel Time Reliability.* UK Department of Transportation, Her Majesty's Stationery Office, London.

Brownstone, D., and K.A. Small (2005). Valuing Time and Reliability: Assessing the Evidence from Road Pricing Demonstrations. *Transportation Research Part A: Policy and Practice*, Vol. 39, pp. 279–293.

Carrion, C., and D. Levinson (2012). Value of Travel Time Reliability: A Review of Current Evidence. *Transportation Research Part A*, Vol. 48, pp. 720–741.

Carrion, C., and D. Levinson (2013). Valuation of Travel Time Reliability from a GPS-Based Experimental Design. *Transportation Research Part C: Emerging Technologies 35,* pp. 305-323.

Cherlow, J.R. (1981). Measuring Values of Travel Time Savings. *Journal of Consumer Research,* Vol. 7, pp. 360–371.

Concas, S., and A. Kolpakov (2009). Synthesis of Research on Value of Time and Value of Reliability. Center for Urban Transportation Research, Tampa, Florida.

Copley, G., and P. Murphy (2002). Understanding and Valuing Journey Time Variability. Publication of: Association for European Transport, United Kingdom.

Devarasetty, P.C., M. Burris, and W.D. Shaw (2012). The Value of Travel Time and Reliability—Evidence from a Stated Preference and Actual Usage. *Transportation Research Part A,* Vol. 46, pp. 1227–1240.

HNTB Corporation (2010). *Atlanta Regional Managed Lane System Plan, Technical Memorandum 6: GDOT Tolling and Traffic Revenue Primer.* < http://www.dot.ga.gov/Projects/studies/managedlanes/Documents/FINALREPORT.pdf > Accessed on September 10, 2013.

Hensher, D.A. (2001). The Valuation of Commuter Travel Time Savings for Car Drivers: Evaluating Alternative Model Specifications. *Transportation*, Vol. 28, No. 2, pp. 101–118.

Lam, T.C., and K.A. Small (2001). The Value of Time and Reliability: Measurement from a Value Pricing Experiment. *Transportation Research Part E: Logistics and Transportation Review*, Vol. 37, pp. 231–251.

Li, Z., D. A. Hensher, and J. M. Rose (2010). Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. *Transportation Research Part E: Logistics and Transportation review*, 46(3), pp. 384-403.

Liu, X., G. Zhang, Y. Lao and Y. Wang (2011). Quantifying the Attractiveness of High-Occupancy Toll Lanes with Traffic Sensor Data Under Various Traffic Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, 2229(1), 102-109.

Netherlands Institute for Transport Policy Analysis (2013). *The Social Value of Shorter and More Reliable Travel Times.* Ministry of Infrastructure and the Environment, Hague, Netherlands.

Oort, C.J. (1969). The Evaluation of Travelling Time. *Journal of Transport Economics and Policy,* Vol. 3, p. 8.

Patil, S., M. Burris, W.D. Shaw, and S. Concas (2011). Variation in the Value of Travel Time Savings and Its Impact on the Benefits of Managed Lanes. *Transportation Planning and Technology* 34.6, pp. 547-567.

Perk, V.A., J.S. DeSalvo, T.A. Rodrigues, N.M. Verzosa, and S.C. Bovino (2011). *Improving Value of Travel Time Savings Estimation for More Effective Transportation Project Evaluation.* Technical Report, Prepared for the Florida Department of Transportation, Center for Urban Transportation Research, Tampa, Florida.

Small, K.A., C. Winston, and J. Yan (2005). Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. *Econometrica*, Vol. 73, pp. 1367–1382.

Small, K.A., C. Winston, J. Yan, N. Baum-Snow, and J.A. Gómez-Ibáñez (2006). Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous

Preferences in Policy Design [with Comments]. *Brookings-Wharton Papers on Urban Affairs*, pp. 53–96.

Small, K.A., R. Noland, X. Chu, and D. Lewis (1999). *NCHRP Report 431: Valuation of Travel Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation.* Transportation Research Board, Washington,

D.C.Tilahun, N., and D. Levinson (2010). A Moment of Time: Reliability in Route Choice Using Stated Preference. *Journal of Intelligent Transportation Systems*, Vol. 14, p. 9.

Tseng, Y., and E. Verhoef (2008). Value of Time by Time of Day: A Stated-Preference Study. *Transportation Research Part B*, Vol. 42, pp. 607–618.

Tseng, Y.Y., E. Verhoef, G. de Jong, M. Kouwenhoven, and T. van der Hoorn (2009). A Pilot Study into the Perception of Unreliability of Travel Times Using In-Depth Interviews. *Journal of Choice Modelling,* Vol. 2, No. 1, pp. 8–28.

Wong, H.K., and J.M. Sussman (1973). Dynamic Travel Time Estimation on Highway Networks. *Transportation Research,* Vol. 7, No. 4, pp. 355–370.

APPENDIX A

SAS CODE

```
/******************************************************************************
*****/

/* Program Name: Katy Freeway|Reliability Study|Danda Thesis*/

/* Date Created: 03/12/2014*/

/* Author: Santosh Rao Danda*/

/* Purpose: To identify trips using raw AVI sensor detections on the Katy
freeway*/

/* */

/* */

/* Inputs: AVI text files, Katy toll data (TransKatyMangdLns2012, V-
tols_KatMangdLns2012)

                Lane closure data, Toll schedule data, distance between sensor
data*/

/* Outputs: Raw2012.sas7bdat (Randomized data) and Apr12trips.sas7bdat*/

/******************************************************************************
*****/



/*PART-1*/



/*Step1.1: Import raw avi data for the whole year into sas dataset

              Only detections on sensors along the freeway are imported */



Data Katy.Data2012;

       infile 'Folder Location\Textfiles2012\*.txt';

       input tagID :$12. Antenna Sensor Readtime :time8. Readdate :mmddyy10.;

       if Sensor = 416 or Sensor = 411 or Sensor = 412 or Sensor = 415 or
Sensor = 413
```

```
        or Sensor = 414 or Sensor = 368 or Sensor = 369 or Sensor = 427 or
Sensor = 449

        or Sensor = 459 or Sensor = 396 or Sensor = 443 or Sensor = 451 or
Sensor = 458

        or Sensor = 442 or Sensor = 466 or Sensor = 468 or Sensor = 469 or
Sensor = 467

        or Sensor = 444 or Sensor = 453 or Sensor = 456 or Sensor = 445 or
Sensor = 440

        or Sensor = 454 or Sensor = 455 or Sensor = 441 or Sensor = 426 or
Sensor = 460

        or Sensor = 425 or Sensor = 6 or Sensor = 5 or Sensor = 271 or Sensor =
272

        or Sensor = 203 or Sensor = 199 or sensor=465 then;

        else delete;

        datetime = dhms(Readdate,0,0,Readtime);

        format datetime datetime22.;

run;


/* Step1.2: Weed out erroneous repetitions in original data*/


proc sql noprint;

        Create Table Katy.Data12 as select distinct * from Katy.data2012;

quit;




/* Step1.3: (i) Randomization of all tag IDs

            (ii) Deleting those tag ids which have been detected only once all
through the month*/




/* Simplify tagids to help merge with toll data*/


Data Katy.data12;

        set Katy.data12(drop = antenna readtime readdate);
```

```
       tagID=substr(tagID,5);   /*Remove first 4 characters of tagid to help
facilitate merging with toll data*/

run;


/*Merge AVI and toll data*/


Data Katy.data12;

       set Katy.data12 Katy.hctra;  /*HCTRA dataset obtained by running Merging
HCTRA and AVI.sas program*/

run;


/*Randomization of tag IDs*/


proc sql noprint;

       Create Table Katy.TagID2 as select TagID, count(*) as count from
Katy.data12 group by TagID;

quit;


proc sort data=Katy.TagID2;

    by tagid;

run;


Data Katy.tagid_f (drop = count);

       set Katy.tagid2;

       if count = 1 then delete;

run;


proc sql noprint;

        select count(*)

                into :mycount

                from Katy.tagid_f;

quit;
```

```
data Katy.RandomID( keep = RandID );

 first = 1E7;

 last = 999999999;

 nobs =  &myCount;

 do RandID=last to first by -1;

   c=ranuni(1234);

   if c<=nobs/(RandID+1-first) then do;

     nobs=nobs-1;

     output;

   end;

 end;

run;


data Katy.tagtoran;

      merge  Katy.tagid_f Katy.RandomID;

run;


proc sort data= Katy.data12;

      by tagid;

run;


/* Step1.4: Delete unwanted tagIDs (Those that only appear once throughout the
year) from the dataset */


Data data12_temp;

      merge  Katy.tagid2 Katy.data12;

      by tagid;

run;


Data Katy.Data12(drop = count);
```

```
        set Data12_temp;

        if count = 1 then delete;

run;


/*Step1.5: Replace original tagid with new randomized tagids in the entire
month dataset */


Data Katy.Raw2012 (drop = tagid);

        merge  Katy.tagtoran Katy.data12;

        by tagid;

run;


/*PART-1 COMPLETE*/


/*PART-2: Identifying trips and adding preliminary trip attributes*/


/*Step 2.1: Selecting data from the month of April 2012*/


Proc sql noprint;

        Create Table Katy.Apr12_2 as select * from Katy.raw2012 where
month(datepart(datetime))=4;

quit;


/*Step 2.2: Selecting distinct records*/


proc sql noprint;

        Create Table Katy.Apr12 as select distinct * from Katy.Apr12_2;

quit;


/*Step 2.3: Deleting sensors outside of the 12 mile stretch of the MLs*/
```

```
Data Katy.Apr12;

	set Katy.Apr12;

	if ((sensor=199) or (sensor=271) or (sensor=272) or (sensor=5) or

		(sensor=6) or (sensor=368) or (sensor=369) or (sensor=413) or

		(sensor=414) or (sensor=415) or (sensor=412) or (sensor=411) or
(sensor=416)) then delete;

run;


Proc sort data= Katy.Apr12;

	by descending randid datetime;

run;


/*Step 2.4: Add corresponding sensors for alternate trips*/


Data Katy.Apr12;

	set Katy.Apr12;

	if sensor=449 then sensor1=427;

	else if sensor=459 then sensor1=396;

	else if sensor=451 then sensor1=443;

	else if sensor=458 then sensor1=442;

	else if sensor=468 then sensor1=466;

	else if sensor=469 then sensor1=467;

	else if sensor=453 then sensor1=444;

	else if sensor=456 then sensor1=445;

	else if sensor=454 then sensor1=440;

	else if sensor=455 then sensor1=441;

	else if sensor=460 then sensor1=425;

	else if sensor=101 then sensor1=465;

	else if sensor=102 then sensor1=465;

	else if sensor=103 then sensor1=443;

	else if sensor=104 then sensor1=443;
```

```
        else if sensor=105 then sensor1=440;
        else if sensor=106 then sensor1=440;
        else if sensor=107 then sensor1=441;
        else if sensor=108 then sensor1=441;
        else if sensor=109 then sensor1=442;
        else if sensor=110 then sensor1=442;
        else if sensor=111 then sensor1=396;
        else if sensor=112 then sensor1=396;
        else if sensor=427 then sensor1=449;
        else if sensor=396 then sensor1=111;
        else if sensor=443 then sensor1=103;
        else if sensor=442 then sensor1=109;
        else if sensor=466 then sensor1=468;
        else if sensor=467 then sensor1=469;
        else if sensor=444 then sensor1=453;
        else if sensor=445 then sensor1=456;
        else if sensor=440 then sensor1=105;
        else if sensor=441 then sensor1=107;
        else if sensor=425 then sensor1=460;
        else if sensor=465 then sensor1=101;
        else sensor1=sensor;
run;

/*Step 2.5: Add trip direction*/

/*Add 1 and 0 for eastbound and westbound direction. */

Data Katy.Apr12;
        set Katy.Apr12;
        if (sensor = 107 or sensor = 108 or sensor = 109 or sensor = 110 or
sensor = 111
```

```
        or sensor = 112 or sensor = 416 or sensor = 415 or sensor = 414 or
sensor = 369

        or sensor = 396 or sensor = 442 or sensor = 467 or sensor = 445 or
sensor = 441

        or sensor = 425 or sensor = 6 or sensor = 272 or sensor = 199 or sensor
= 459

        or sensor = 458 or sensor = 469 or sensor = 456 or sensor = 455 or
sensor = 460)

        then direction = 1;

        else direction = 0;

run;


/*Step 2.6: Add sensor type*/


Data Katy.Apr12;

        set Katy.Apr12;

        if (sensor = 459 or sensor = 458 or sensor = 469 or sensor = 456 or
sensor = 455

        or sensor = 460 or sensor = 449 or sensor = 451 or sensor = 468 or
sensor = 453

        or sensor = 454) then location = 'M';

        else if (sensor = 101 or sensor = 103 or sensor = 105 or sensor = 107 or
sensor = 109

        or sensor = 111) then location = 'T';

        else if (sensor = 102 or sensor = 104 or sensor = 106 or sensor = 108 or
sensor = 110

        or sensor = 112) then location = 'H';

        else if (sensor = 396 or sensor = 442 or sensor = 467 or sensor = 445 or
sensor = 441

        or sensor = 425 or sensor = 440 or sensor = 444 or sensor = 466 or
sensor = 443

        or sensor = 427 or sensor = 465 or sensor=426) then location = 'G';

        else location  = 'O';

run;
```

```
/*Step 2.7: Add sensor type for alternate trip sensors*/


Data Katy.Apr12;

       set Katy.Apr12;

       if (sensor1 = 459 or sensor1 = 458 or sensor1 = 469 or sensor1 = 456 or
sensor1 = 455

       or sensor1 = 460 or sensor1 = 449 or sensor1 = 451 or sensor1 = 468 or
sensor1 = 453

       or sensor1 = 454) then location1 = 'M';

       else if (sensor1 = 101 or sensor1 = 103 or sensor1 = 105 or sensor1 =
107 or sensor1 = 109

       or sensor1 = 111) then location1 = 'T';

       else if (sensor1 = 102 or sensor1 = 104 or sensor1 = 106 or sensor1 =
108 or sensor1 = 110

       or sensor1 = 112) then location1 = 'H';

       else if (sensor1 = 396 or sensor1 = 442 or sensor1 = 467 or sensor1 =
445 or sensor1 = 441

       or sensor1 = 425 or sensor1 = 440 or sensor1 = 444 or sensor1 = 466 or
sensor1 = 443

       or sensor1 = 427 or sensor1 = 465 or sensor1=426) then location1 = 'G';

       else location1 = 'O';

run;


/*Step 2.8: Identify trip start*/


Data Katy.Apr12;

       set Katy.Apr12;

    start = 0;

    end = 0;

    timediff = datetime-lag(datetime);

    if randid ne lag(randid) then start=1;

       if (timediff > hms(0,10,0)) then start=1;

       if (direction ne lag(direction)) then start=1;
```

```
run;

/*Step 2.9: Identify trip ends*/

Data Katy.Apr12;
    set Katy.Apr12;
    obs= _N_;
run;

Proc sort data= Katy.Apr12;
       by descending obs;
run;

Data Katy.Apr12;
       set Katy.Apr12;
       if lag(start)=1 then end=1;
run;

Proc sort data= Katy.Apr12;
       by obs;
run;

/*Step 2.10: Delete trips which are only detected at one sensor*/

Data Katy.Apr12;
    set Katy.Apr12;
    if ((start=1) and (end=1)) then delete;
run;

/*Step 2.11: Create variable with all sensors for a trip*/
```

```
data Katy.Apr12;
	set Katy.Apr12;

	length Allsensor Allsensor1 $100;
	retain INTER INTER1;

	if Start = 1 then
		do;
			Allsensor = catx(",", sensor);
			INTER = Allsensor;
			Allsensor1 = catx(",", sensor1);
			INTER1 = Allsensor1;
		end;
	else
		do;
			Allsensor = catx(",", INTER, sensor );
			INTER = Allsensor;
			Allsensor1 = catx(",", INTER1, sensor1 );
			INTER1 = Allsensor1;
		end;
	drop INTER INTER1;
run;

/*Step 2.12: Create variable with all sensor types for a trip*/

Data Katy.Apr12;
	length Weave Weave1 $50;
	length INTER INTER1 $50;
	set Katy.Apr12;
```

```
        retain INTER INTER1;

        if Start = 1 then
              do;
                    Weave = catx(" ", location);
                    INTER = Weave;
                    Weave1 = catx(" ", location1);
                    INTER1 = Weave1;
              end;
        else
              do;
                    Weave = catx(" ", INTER, location);
                    INTER = Weave;
                    Weave1 = catx(" ", INTER1, location1);
                    INTER1 = Weave1;
              end;
        drop INTER INTER1;
run;

/* Step 2.13: Identifying trip start sensor, trip start time, trip end sensor
and trip end time */

Data Katy.Apr12;
        set Katy.Apr12;
        retain Inter Inter2;
     if start=1 then
         do;
                    Start_Sensor= Sensor;
               Inter=Start_Sensor;
             Starttime=datetime;
               Inter2= Starttime;
```

```
              Time=0;
              End_Sensor=0;
            Endtime=0;
          end;

      if end=1 then
         do;
              Start_Sensor= Inter;
              Starttime = Inter2;
              End_Sensor=Sensor;
              Endtime=datetime;
              Time= datetime-Inter2;
         end;

       if ((start=0) and (end=0)) then
          do;
              Start_Sensor= Inter;
              Inter= Start_Sensor;
              Starttime=Inter2;
              Inter2=Starttime;
              Time=0;
              End_Sensor=0;
              Endtime=0;
          end;
       drop inter inter2;
  run;

/*Step 2.14: Identifying start and end sensors for the dummy trip*/

  Data Katy.Apr12;
```

```
      set Katy.Apr12;
      retain Inter;
    if start=1 then
        do;
                    Start_Sensor1= Sensor1;
                Inter=Start_Sensor1;
                End_Sensor1=0;
          end;

     if end=1 then
          do;
               Start_Sensor1= Inter;
               End_Sensor1=Sensor1;
          end;

      if ((start=0) and (end=0)) then
           do;
               Start_Sensor1= Inter;
               Inter= Start_Sensor1;
               End_Sensor1=0;
           end;
      drop inter;
  run;

/*Step 2.15: Inputting trip end variables in all the trip observations*/

proc sort data= Katy.Apr12;
    by descending obs;
run;
```

```
Data Katy.Apr12;
	set Katy.Apr12;
	retain Inter Inter2 Inter3;

	if end=1 then
		do;
			Inter= Time;
			Inter2= End_Sensor;
			Inter3= Endtime;
	end;

	if (((start=0) and (end=0)) or (Start=1)) then
	    do;
			Time= Inter;
			Inter= Time;
			End_Sensor= Inter2;
			Inter2= End_Sensor;
			Endtime= Inter3;
			Inter3= Endtime;
		end;
	drop Inter Inter2 Inter3;
run;

Data Katy.Apr12;
	set Katy.Apr12;
	retain Inter2;

	if end=1 then
		do;
			Inter2= End_Sensor1;
```

```
            end;

        if (((start=0) and (end=0)) or (Start=1)) then
            do;
                        End_Sensor1= Inter2;
                        Inter2= End_Sensor1;
                end;
        drop Inter2;
run;

proc sort data= Katy.Apr12;
        by obs;
run;

/* PART-2 COMPLETE */

/*PART-3: Inputting lane closure information*/

/*Step 3.1: Importing lane closure data file*/
proc import
        datafile = "C:\Users\s-danda\Dropbox\My
Codes\Excel_files\Laneclosure2012_brad.xlsx"
    out = Katy.laneclosure
        dbms = xlsx;
        sheet = "2012_I-10Katy_AllIncidents";
        getnames = yes;
run;

Proc sort data=Katy.laneclosure;
```

```
        by cleared_date_time;

run;


data Katy.laneclosure;

        set Katy.laneclosure;

                incidentstart=detection_date_time;

                incidentend=cleared_date_time;

        if incidentend=253717747199 then delete; /*Deleting improper records 36
incidents deleted.*/

run;


data Katy.Apr12;

        set Katy.Apr12;

        detectiontime=datetime;

run;


/*Step 3.2: Merging Trips data and lane closure data based on incident time and
nearest sensors*/


proc sql noprint;

create table Katy.xyz as

select A.*, B.mainlanes_blocked, B.frontage_lanes_blocked,
B.ramp_lanes_blocked, B.HOV_lanes_blocked,

B.shoulder_lanes_blocked, B.detection_date_time,B.cleared_date_time

from Katy.Apr12 as A left join Katy.laneclosure as B

on (B.incidentstart < A.detectiontime < B.incidentend) and (A.sensor=
B.nearest_avi_sensor);

quit;


data Katy.Apr12;

        set Katy.xyz;

run;
```

```
proc sort data= Katy.Apr12;
	by obs;
run;


/*Step 3.3: Making sure the highest number of lanes blocked are recorded for
each trip*/


Data Katy.Apr12;
	set Katy.Apr12;
	retain mainlanes_blocked frontage_lanes_blocked ramp_lanes_blocked
HOV_lanes_blocked Shoulder_lanes_blocked;
	inter1= lag1(mainlanes_blocked);
	inter2= lag1(frontage_lanes_blocked);
	inter3= lag1(ramp_lanes_blocked);
	inter4= lag1(HOV_lanes_blocked);
	inter5= lag1(Shoulder_lanes_blocked);
	if ((start=0) and (end=1)) or ((start=0) and (end=0)) and (inter1 ne .
or inter2 ne . or inter3 ne . or inter4 ne . or inter5 ne .)
	then do;
		mainlanes_blocked= max(mainlanes_blocked, inter1);
		frontage_lanes_blocked= max(frontage_lanes_blocked, inter2);
		ramp_lanes_blocked= max(ramp_lanes_blocked, inter3);
		HOV_lanes_blocked= max(HOV_lanes_blocked, inter4);
		Shoulder_lanes_blocked= max(Shoulder_lanes_blocked, inter5);

	end;
run;


data Katy.Apr12;
	set Katy.Apr12;
	drop inter1 inter2 inter3 inter4 inter5  detection_date_time
cleared_date_time ;
```

```
run;

Data Katy.Apr12;
	set Katy.Apr12;
	if ((randid=lag(randid)) and (sensor=lag(sensor)) and
(datetime=lag(datetime))) then delete;
run;

/*PART-3 COMPLETE*/

/*PART-4: Assigning trip lengths*/

/*Step 4.1: Inputting total trip segments*/

Data Katy.Apr12;
	set Katy.Apr12;
	if start=1 then
		do;
			segmentno=0;
		end;
	if start=0 then
		do;
			segmentno+1;
		end;
run;

/*Step 4.2: Sort trip length data file*/

Proc sort data= Katy.distance;
	by sensor;
run;
```

```
Proc sort data= Katy.Apr12;
      by sensor;
run;

/*Step 4.3: Merge trips data and trip length data based on sensor coordinates*/

data Katy.Apr12;
      merge  Katy.Apr12 Katy.distance;
      by sensor;
run;

Data Katy.Apr12;
      set Katy.Apr12;
      if datetime=. then delete;
run;

/*Step 4.4: Check to see for negative lengths*/

Data Katy.Apr12;
      set Katy.Apr12;
      retain inter;
      if (start=1) then
            do;
                  Inter = coord;
                  seglen=0;
      end;

      if (start=0) then
            do;
```

```
                    seglen=(coord-inter);
                    Inter= coord;
                    if seglen<0 then flag=1;
                end;
        drop inter;
run;

proc sort data= Katy.Apr12;
    by descending obs;
run;

Data Katy.Apr12;
        set Katy.Apr12;
        retain Inter Inter1;

        if end=1 then
                do;
                    Inter= Weave;
                    Inter1= Weave1;
                end;

        if (((start=0) and (end=0)) or (Start=1)) then
            do;
                    Weave= Inter;
                    Inter= Weave;
                    Weave1= Inter1;
                    Inter1= Weave1;
                end;
        drop Inter Inter1;
run;
```

```
proc sort data= Katy.Apr12;
       by obs;
run;

Data Katy.Apr12;
       set Katy.Apr12;
       retain inter;
       if ((start=0) and (inter=1)) then flag=1;
       inter=flag;
       drop inter;
run;

proc sort data= Katy.Apr12;
    by descending obs;
run;

Data Katy.Apr12;
       set Katy.Apr12;
       retain inter;
       if end=0 then
              do;
                     flag=inter;
              end;
       inter=flag;
       drop inter;
run;

proc sort data= Katy.Apr12;
    by obs;
```

```
run;

Data Katy.Apr12;
       set Katy.Apr12;
       if flag=1 then delete;
       drop flag;
run;

/*Step 4.5: Summing up segment lengths to get trip lengths*/

Data Katy.Apr12;
       set Katy.Apr12;
       retain inter disttravel;

       if (start=1) then
              do;
                     length = 0;
                     Inter = coord;
                     disttravel=0;
       end;

       if (start=0) then
              do;
                     length= (coord-inter)+disttravel;
                     Inter= coord;
                     disttravel=length;
       end;
       drop inter disttravel;
run;
```

```
/*Step 4.6: Calculating trip lengths for alternate trips*/

Data Katy.distance;
	set Katy.distance(rename=(sensor= sensor1 coord=coord1));
run;

proc sort data= Katy.Apr12;
	by sensor1;
run;

data Katy.Apr12;
	merge  Katy.Apr12 Katy.distance;
	by sensor1;
run;

proc sort data= Katy.Apr12;
	by obs;
run;

data Katy.Apr12;
	set Katy.Apr12;
	retain inter disttravel;

	if (start=1) then
		do;
			length1 = 0;
			Inter = coord1;
			disttravel=0;
	end;
```

```
        if (start=0) then
              do;
                     length1= (coord1-inter)+disttravel;
                     Inter= coord1;
                     disttravel=length1;
        end;
        drop inter disttravel;
run;

Data Katy.distance;
        set Katy.distance(rename=(sensor1= sensor coord1=coord));
run;

/*Step 4.7: Additional step to delete data pertaining to sensor that may not be
present in the data*/

Data Katy.Apr12;
        set Katy.Apr12;
        if datetime=. then delete;
run;

/*PART 4 COMPLETE*/

/*PART 5: Assign Toll*/

/*Step 5.1: Importing toll schedules*/

proc import
        datafile = "Folder Location\KatyTollPrice_preSep82012.xls"
    out = EB_Seg1
        dbms = xls;
```

```
		sheet = "EB_Seg1";
		getnames = yes;
run;

proc import
		datafile = "Folder Location\KatyTollPrice_preSep82012.xls"
	out = EB_Seg23
		dbms = xls;
		sheet = "EB_Seg23";
		getnames = yes;
run;

proc import
		datafile = "Folder Location\KatyTollPrice_preSep82012.xls"
	out = WB_Seg1
		dbms = xls;
		sheet = "WB_Seg1";
		getnames = yes;
run;

proc import
		datafile = "Folder Location\KatyTollPrice_preSep82012.xls"
	out = WB_Seg23
		dbms = xls;
		sheet = "WB_Seg23";
		getnames = yes;
run;

/********** ******************** ******************** ***********/
/* Step 5.2: Creating functions to save processing time */
```

```
/********** ******************** ******************** ***********/

proc fcmp outlib = work.FCMP_TollPrice.fcns;

       function obtainToll(fileName $, hour, weekday, location $);

              array tollPrice[24, 7] / nosymbols;

              rc = read_array(fileName,tollPrice,'B','C','D','E','F','G','H');

           if location = 'T' then return(tollPrice[hour,weekday]);

              if location = 'H' then do;

                     if (weekday > 1 and weekday < 7) and ((hour >= 15 and hour
< 21) or (hour >= 6 and hour < 12)) then return(0.0);

                        else return(tollPrice[hour,weekday]);

                     end;

              return( 0.0 );

       endsub;

run;

data Katy.Apr12;

       set Katy.Apr12;

       hour = hour( datetime );

       date = datepart( datetime );

run;

options cmplib = work.FCMP_TollPrice;

/*Step 5.3: Assigning tolls to trips based on detections*/

Data Katy.Apr12;

       set Katy.Apr12;
```

```
	select( sensor );

		when ( 101, 102 ) Toll =
obtainToll('work.Eb_seg1',input(hour,best32.)+1, weekday(date),location);
/*note hour starts from 0*/

		when ( 103, 104, 105, 106 ) Toll = obtainToll('work.Eb_seg23',
input(hour,best32.)+1, weekday(date),location);

	    when ( 111, 112 ) Toll = obtainToll('work.Wb_seg1',
input(hour,best32.)+1, weekday(date),location);

		when ( 107, 108, 109, 110 ) Toll = obtainToll('work.Wb_seg23',
input(hour,best32.)+1, weekday(date),location);

		otherwise Toll = 0.0;

	end;

	/* drop hour date; */

run;



/*Step 5.4: Assigning tolls to alternate trips*/



Data Katy.Apr12;

	set Katy.Apr12;

	select( sensor1 );

		when ( 101, 102 ) Toll1 =
obtainToll('work.Eb_seg1',input(hour,best32.)+1, weekday(date),location1);
/*note hour starts from 0*/

		when ( 103, 104, 105, 106 ) Toll1 = obtainToll('work.Eb_seg23',
input(hour,best32.)+1, weekday(date),location1);

	    when ( 111, 112 ) Toll1 = obtainToll('work.Wb_seg1',
input(hour,best32.)+1, weekday(date),location1);

		when ( 107, 108, 109, 110 ) Toll1 = obtainToll('work.Wb_seg23',
input(hour,best32.)+1, weekday(date),location1);

		otherwise Toll1 = 0.0;

	end;

	/* drop hour date; */

run;



/* Step 5.5: Delete functions */
```

```
proc fcmp outlib = work.FCMP_TollPrice.fcns;
  deletefunc obtainToll;
quit;

/* Step 5.6: Summing up the tolls along a trip*/

data Katy.Apr12;
	set Katy.Apr12;
	retain inter inter1;
	if (start=1) then

		do;
			Totaltoll = toll;
			Inter = Totaltoll;
			Totaltoll1 = toll1;
			Inter1 = Totaltoll1;
		end;

	if (start=0) then
		do;
			Totaltoll= toll + Inter;
			Inter= Totaltoll;
			Totaltoll1= toll1 + Inter1;
			Inter1= Totaltoll1;
		end;
	drop inter inter1;
run;

/*PART-5 COMPLETE*/
```

```
/*PART-6: Calculating Travel time and travel time standard deviation*/

/*Step 6.1: Computing average travel times and travel time std for
10 min intervals per segment*/

data Katy.Apr12;
       set Katy.Apr12;
       startdate=datepart(datetime);
       starttime_sec = timepart(starttime);
       marker10= floor(starttime_sec/600);
       starttime_10min=marker10*600;
run;

Data Katy.Apr12;
       set Katy.Apr12;
       retain sensor sensor1;
       startsensor1=lag(sensor);
       endsensor1=sensor;
       startsensor11=lag(sensor1);
       endsensor11=sensor1;
       count=1;
       if start=1 then timediff=.;
run;

Data abc1;
       set Katy.Apr12;
       if timediff=. then delete;
run;

Proc sql noprint;
```

```
        Create Table Katy.Apr12_segment_stddev

        as select startdate, startsensor1, endsensor1, starttime_10min,

        sum(count) as Count_10_segment, avg(timediff) as TT_ave_10min_segment,

        std(timediff) as TT_std_10min_segment

        from abc1

        group by startdate, startsensor1, endsensor1, starttime_10min;

quit;


/*Step 6.2: Merging segment travel time and standard deviation data with trip
segment data data*/


proc sort data= Katy.Apr12;

        by startdate startsensor1 endsensor1 starttime_10min;

run;


Data Katy.Apr12;

        merge Katy.Apr12 Katy.Apr12_segment_stddev;

        by startdate startsensor1 endsensor1 starttime_10min;

run;


proc sort data= Katy.Apr12;

        by descending randid datetime;

run;


Data Katy.Apr12;

        set Katy.Apr12;

        if randid=. then delete;

run;


/*Step 6.3: Merging segment travel time and std data with alternate trips*/
```

```
Data Katy.Apr12_segment_stddev; /*Renaming variable to enable merging according to dummy trip*/

	set Katy.Apr12_segment_stddev(rename=(startsensor1=startsensor11 endsensor1=endsensor11 Count_10_segment=Count_10_segment1

		TT_ave_10min_segment=TT_ave_10min_segment1 TT_std_10min_segment=TT_std_10min_segment1 ));

run;


proc sort data= Katy.Apr12;

	by startdate startsensor11 endsensor11 starttime_10min;

run;


Data Katy.Apr12;

	merge Katy.Apr12 Katy.Apr12_segment_stddev;

	by startdate startsensor11 endsensor11 starttime_10min;

run;


proc sort data= Katy.Apr12;

	by descending randid datetime;

run;


Data Katy.Apr12;

	set Katy.Apr12;

	if randid=. then delete;

run;


Data Katy.Apr12_segment_stddev; /*Renaming variables back to original name*/

	set Katy.Apr12_segment_stddev(rename=(startsensor11=startsensor1 endsensor11=endsensor1 Count_10_segment1=Count_10_segment

		TT_ave_10min_segment1=TT_ave_10min_segment TT_std_10min_segment1=TT_std_10min_segment));

run;
```

```
/* Step 6.3: Assigning segment lengths to all segments */

Data Katy.Apr12;
	set Katy.Apr12;
	retain inter inter1;

	if (start=1) then
		do;
			segmentlength = 0;
			Inter = length;
			segmentlength1 = 0;
			Inter1 = length1;
	end;

	if (start=0) then
		do;
			segmentlength= (length-Inter);
			Inter= length;
			segmentlength1= (length1-Inter1);
			Inter1= length1;
	end;
	drop inter inter1;
run;

/*Step 6.4: Categorizing trips as peak, shoulder or off-peak based on time of
trip*/

Data Katy.Apr12;
	set Katy.Apr12;
	weekday=weekday(datepart(starttime));
	hour=hour(starttime);
```

```
        peak1=0;

        peak2=0;

        if ((weekday ne 1) and (weekday ne 7)) then

               do;

                      if ((direction=0) and (7<=hour<=8)) or ((direction=1) and
(16<=hour<=17)) then peak1=1;

                      if ((direction=0) and (6<=hour<=9)) or ((direction=1) and
(15<=hour<=18)) then peak2=1;

               end;

        drop weekday hour;

run;


/*Step 6.5: Eliminating trips that are detected at toll as well as GPL
sensors*/


data Katy.Apr12;

        set Katy.Apr12(drop = startdate starttime_sec marker10 starttime_10min);

        g= countc(Weave, 'G');

        m= countc(Weave,'M');

     o= countc(Weave,'O');

        h= countc(Weave,'H');

        t= countc(Weave,'T');

        if (g=m=h=t=0) then delete;

        if (((m>0) or (h>0) or (t>0)) and (g>0)) then delete;

        if ((m>0) or (h>0) or (t>0)) then type=1;

        if (g>0) then type=2;

        drop g m o h t;

run;


/*Step 6.6: Approximating segment travel times using average speed data*/


Data Katy.Apr12;
```

```
set Katy.Apr12;
if ((start ne 1) and (TT_ave_10min_segment=.)) then
	do;
		if type=1 then
			do;
				if ((peak1=1) and (peak2=1)) then
TT_ave_10min_segment= (segmentlength/53.1)*3600;
				else if ((peak1=0) and (peak2=1)) then
TT_ave_10min_segment= (segmentlength/61.8)*3600;
				else if ((peak1=0) and (peak2=0)) then
TT_ave_10min_segment= (segmentlength/70.9)*3600;
			end;

		else if type=2 then
			do;
				if ((peak1=1) and (peak2=1)) then
TT_ave_10min_segment= (segmentlength/40.8)*3600;
				else if ((peak1=0) and (peak2=1)) then
TT_ave_10min_segment= (segmentlength/55.1)*3600;
				else if ((peak1=0) and (peak2=0)) then
TT_ave_10min_segment= (segmentlength/65.3)*3600;
			end;
	end;
if ((start ne 1) and (TT_ave_10min_segment1=.)) then
	do;
		if type=2 then
			do;
				if ((peak1=1) and (peak2=1)) then
TT_ave_10min_segment1= (segmentlength1/53.1)*3600;
				else if ((peak1=0) and (peak2=1)) then
TT_ave_10min_segment1= (segmentlength1/61.8)*3600;
				else if ((peak1=0) and (peak2=0)) then
TT_ave_10min_segment1= (segmentlength1/70.9)*3600;
			end;
```

```
                    else if type=1 then

                          do;

                                if ((peak1=1) and (peak2=1)) then
TT_ave_10min_segment1= (segmentlength1/40.8)*3600;

                                else if ((peak1=0) and (peak2=1)) then
TT_ave_10min_segment1= (segmentlength1/55.1)*3600;

                                else if ((peak1=0) and (peak2=0)) then
TT_ave_10min_segment1= (segmentlength1/65.3)*3600;

                          end;

              end;

run;


/*Step 6.7: Approximating std=0 if data available within a 10 min interval
between sensor pairs*/


Data Katy.Apr12;

       set Katy.Apr12;

       if ((start ne 1) and (TT_std_10min_segment=.)) then

              do;

                     TT_std_10min_segment= 0;

              end;

       if ((start ne 1) and (TT_std_10min_segment1=.)) then

              do;

                     TT_std_10min_segment1= 0;

              end;

run;


Data Katy.Apr12;

       set Katy.Apr12;

       drop peak1 peak2 type;

run;
```

```
/* Step 6.8: Calculating standard deviation by summing up all the std of
individual segments*/


Data Katy.Apr12;

	set Katy.Apr12;

	if start=1 then

		do;

			calcstd = 0;

			calcstd1= 0;

			calcave= 0;

			calcave1= 0;

		end;


	if start=0 then

		do;

			calcstd + TT_std_10min_segment;

			calcstd1 + TT_std_10min_segment1;

			calcave + TT_ave_10min_segment;

			calcave1 + TT_ave_10min_segment1;

		end;

run;


/*PART-6 COMPLETE*/


/*PART-7: Creating new data set with complete trips */


Data Katy.Aprend;

	set Katy.Apr12(drop = startsensor1 endsensor1 startsensor11 endsensor11
Count_10_segment Count_10_segment1

				TT_std_10min_segment TT_std_10min_segment1
TT_ave_10min_segment TT_ave_10min_segment1
```

```
                              segmentlength segmentlength1 /* startdate
starttime_sec marker10 starttime_10min */);

       if end=1;

run;


/*PART-7 COMPLETE*/


/*PART-8: Calculating travel times and travel time std for the entire trip*/


/*Step 8.1: Computing average travel times and travel time std for

10 min intervals between trip start and end sensor */


data Katy.Aprend;

       set Katy.Aprend;

       startdate=datepart(datetime);

       starttime_sec = timepart(starttime);

       marker10= floor(starttime_sec/600);

       starttime_10min=marker10*600;

run;


Proc sql noprint;

       Create Table Katy.Apr12stddev

       as select startdate, start_sensor, end_sensor, starttime_10min,
sum(count) as Count_10, avg(time) as TT_ave_10min, std(time) as TT_std_10min

       from Katy.Aprend

       group by startdate, start_sensor, end_sensor, starttime_10min;

quit;


/*Step 8.2: Merging trips data with travel time and standard deviation data per
10 min interval */


proc sort data= Katy.Aprend;
```

```
	by startdate start_sensor end_sensor starttime_10min;
run;

Data Katy.Aprend;
	merge Katy.Aprend Katy.Apr12stddev;
	by startdate start_sensor end_sensor starttime_10min;
run;

proc sort data= Katy.Aprend;
	by descending randid datetime;
run;

Data Katy.Aprend;
	set Katy.Aprend;
	if randid=. then delete;
run;

Data Katy.Apr12stddev; /*Renaming variable to enable merging according to dummy trip*/
	set Katy.Apr12stddev(rename=(start_sensor=start_sensor1 end_sensor=end_sensor1 Count_10=Count_101
		TT_ave_10min=TT_ave_10min1 TT_std_10min=TT_std_10min1 ));
run;

proc sort data= Katy.Aprend;
	by startdate start_sensor1 end_sensor1 starttime_10min;
run;

Data Katy.Aprend;
	merge Katy.Aprend Katy.Apr12stddev;
	by startdate start_sensor1 end_sensor1 starttime_10min;
```

```
run;

proc sort data= Katy.Aprend;
	by descending randid datetime;
run;

Data Katy.Aprend;			/*Deleting observations that do not have corresponding observations from Apr12 dataset*/
	set Katy.Aprend;
	if randid=. then delete;
run;

Data Katy.Apr12stddev; /*Renaming variables back to original name*/
	set Katy.Apr12stddev(rename=(start_sensor1=start_sensor end_sensor1=end_sensor Count_101=Count_10
		TT_ave_10min1=TT_ave_10min TT_std_10min1=TT_std_10min));
run;

/*Step 8.3: Deleting trip observations where the trip or the dummy trip goes through the managed lanes
  but is not detected at the toll sensors*/

Data Katy.Aprend;
	set Katy.Aprend;
	if totaltoll=0 and totaltoll1=0 then delete;
run;

/*Step 8.4: Assigning time and std*/

Data Katy.Aprend;
	set Katy.Aprend;
	if count_10>2 then
```

```
            do;
                std=TT_std_10min;
                time2=TT_ave_10min;
            end;
        else do;
                std=6.37 + 0.48*calcstd+2.20*segmentno;
                time2=calcave;
            end;
        if count_101>2 then
            do;
                time1=TT_ave_10min1;
                std1=TT_std_10min1;
            end;
        else do;
                time1=calcave1;
                std1=6.37+ 0.48*calcstd1+2.20*segmentno;
            end;
        if calcstd=0 then std=0;
        if calcstd1=0 then std1=0;
run;

/*PART 8 COMPLETE*/

/*PART-9: Adding other trip details and attributes*/

/*Step 9.1: Assigning trip number to each trip of a specific randid*/

Data Katy.Aprend;
        set Katy.Aprend;
        if randid ne lag(randid) then tripno=1;
```

```
        if randid=lag(randid) then tripno+1;
run;

/*Step 9.2: Assigning total trips identified per randid*/

Proc sort data=Katy.Aprend;
        by descending obs;
run;

Data Katy.Aprend;
        set Katy.Aprend;
        retain inter;
        if (randid ne lag(randid)) then totaltrips=tripno;
        if (randid=lag(randid)) then totaltrips=inter;
        inter=totaltrips;
        drop inter;
run;

Proc sort data=Katy.Aprend;
        by obs;
run;

/*Step 9.3: Determining trips made during the peak hours and the trips
made during the peak+shoulder hours*/

Data Katy.Aprend;
        set Katy.Aprend;
        cov=std/time2;
        cov1=std1/time1;
        weekday=weekday(datepart(starttime));
```

```
	hour=hour(starttime);

	peak1=0;

	peak2=0;

	if ((weekday ne 1) and (weekday ne 7)) then

		do;

			if ((direction=0) and (7<=hour<=8)) or ((direction=1) and
(16<=hour<=17)) then peak1=1;

			if ((direction=0) and (6<=hour<=9)) or ((direction=1) and
(15<=hour<=18)) then peak2=1;

		end;

	drop weekday hour;

run;


/*Step 9.4: Eliminating trips with 0 length*/


Data Katy.Aprend;

	set Katy.Aprend;

	if ((length=0) or (length1=0)) then delete;

run;


/*Step 9.5: Additional step added to eliminate detection sequence such as 101
followed by 102*/


Data Katy.Aprend;

	set Katy.Aprend;

	if time=0 or time1=0 then delete;

run;


/*PART-9 COMPLETE*/


/*PART 10: Reorganizing data for choice modeling */
```

```
/*Step 10.1: Assigning trip id to each of the trips*/

Data Katy.Aprend;
	set Katy.Aprend;
	id+1;
run;

/*Step 10.2: Creating a separate record for each trip and alternate trip for modeling*/

Data Katy.Apr12trips (keep= randid start_sensor end_sensor starttime endtime
				weave obs allsensor time segmentno length
Totaltoll count_10
				std tripno totaltrips lanechoice peak1 peak2
id
				mainlanes_blocked frontage_lanes_blocked
ramp_lanes_blocked
				hov_lanes_blocked shoulder_lanes_blocked
cov);
	set Katy.Aprend;
	lanechoice=1;
	output;
	lanechoice=0;
	start_sensor=start_sensor1;
	end_sensor=end_sensor1;
	weave=weave1;
	allsensor=allsensor1;
	time=time1;
	length=length1;
	totaltoll=totaltoll1;
	count_10=count_101;
	std=std1;
	cov=cov1;
```

```
        output;
run;

/*Step 10.3: Assigning trip type. 1 for ML and 2 for GPL*/

Data Katy.Apr12trips;
        retain obs id randid tripno totaltrips type lanechoice time std
Totaltoll
                count_10 length segmentno start_sensor end_sensor weave
allsensor
                peak1 peak2 starttime endtime mainlanes_blocked
                frontage_lanes_blocked ramp_lanes_blocked hov_lanes_blocked
                shoulder_lanes_blocked cov;
        set Katy.Apr12trips;
        if totaltoll>0 then type=1;
        else type=2;
run;

Proc sort data=Katy.Apr12trips;
        by obs type;
run;

/*Step 10.4: Add dummy values for ASC_ML and ASC_GPL*/

Data Katy.Apr12trips;
        set Katy.Apr12trips;
        MLdum=0;
        GPLdum=0;
        if type=1 then
                do;
                        MLdum=1;
```

```
                GPLdum=0;
            end;
        if type=2 then
            do;
                MLdum=0;
                GPLdum=1;
            end;
run;

/*Step 10.5: Converting time and std into minutes*/

Data Katy.Apr12trips;
        set Katy.Apr12trips;
        time=time/60;
        std=std/60;
run;

/*Step 10.6: Adding interaction terms,squared terms and relative std*/

Data Katy.Apr12trips;
        set Katy.Apr12trips;
        std_len=std/length;
        intmultiply=time*totaltoll;
        intdivide=totaltoll/time;
        totaltoll2=totaltoll*totaltoll;
        time2=time*time;
        std2=std*std;
        relstd=std/time;
run;
```

```
/*Step 10.7: Dividing data into peak, off-peak and shoulder trips*/

data Katy.Apr12trips_peak Katy.Apr12trips_off Katy.Apr12trips_shou;
	set Katy.Apr12trips;
	if peak1=1 then output Katy.Apr12trips_peak;
	if peak1=0 and peak2=1 then output Katy.Apr12trips_shou;
	if peak1=0 and peak2=0 then output Katy.Apr12trips_off;
run;

/*Step 10.8: Categorizing trips based on length of trip*/

Data Katy.Apr12trips;
	set Katy.Apr12trips;
	if type=1 then
		do;
			if length<3.5 then category=1;
			if (length>=3.5 and length<=8) then category=2;
			if length>8 then category=3;
		end;
run;

Data Katy.Apr12trips;
	set Katy.Apr12trips;
	retain inter;
	if type=2 then
		do;
			category=inter;
		end;
	inter=category;
run;
```

```
Data Katy.Apr12trips_short Katy.Apr12trips_medium Katy.Apr12trips_long;
	set Katy.Apr12trips;
	if category=1 then output Katy.Apr12trips_short;
	if category=2 then output Katy.Apr12trips_medium;
	if category=3 then output Katy.Apr12trips_long;
run;

/*CODE COMPLETE*/

/*CODE APPENDIX A: Approximating Average Speeds*/

Data Katy.SpeedApprox;
	set Katy.Apr12trips;
	if lanechoice=1;
	speed=(length/time)*60;
run;

Proc sql noprint;
	Create Table Katy.Speed
	as select type, peak1, peak2, avg(speed) as sp
	from Katy.SpeedApprox
	group by type, peak1, peak2;
quit;

/*APPENDIX A COMPLETE*/

/*CODE APPENDIX B: Developing regression model for std*/

proc sql noprint;
```

```
       Create Table Katy.Regression as select count_10, tt_std_10min, calcstd,
segmentno from Katy.Aprend where count_10>4;

quit;


/*Generating regression parameters relating estimated and actual standard
deviation*/


 proc reg data=Katy.Regression;

      model TT_std_10min = calcstd segmentno;

 run;


/*APPENDIX B COMPLETE*/


/*CODE APPENDIX C: Sample discrete choice model code*/


proc mdc data=katy.Apr12trips;

       model lanechoice = MLdum time std totaltoll /

       type=clogit

        nchoice=2;

    id id;

run;


/*APPENDIX C COMPLETE*/
```