ABSTRACT

Title of Document: Using Traditional Household Survey and GPS Data for Advanced Travel Behavior and Emission Analysis

Xiaojie Cong, Master of Science, 2012

Directed By: Professor Lei Zhang
Department of Civil and Environmental Engineering

National and regional travel survey data have been widely collected in recent years. This thesis research employs National Household Travel Survey (NHTS) and Maryland GPS survey data sources to accomplish the following two objectives. The primary goal of the study is to assess how residential density, employment density, land use mix, and average block size measured at both the residential locations and at the activity space level influence vehicle miles travelled (VMT) with the Maryland GPS survey data. The secondary goal of the project is to examine the impact of time of day, day of week, trip purpose, vehicle type, gas price on vehicle soak time distributions with the 2009 NHTS data. Econometric models with panel data and Generalized Gamma techniques are developed for the impact analysis.

Using Traditional Household Survey and GPS Data for Advanced Travel Behavior and Emission Analysis

By

XIAOJIE CONG

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2012

Advisory Committee:
Professor Lei Zhang, Chair/Advisor
Professor Paul Schonfeld
Professor Cinzia Cirillo

Dedication

To my parents and my sister for their love.

Acknowledgements

First, I would like to express my sincere appreciation and gratitude to my advisor Dr. Lei Zhang for offering me the opportunity to study at the University of Maryland. I have gained not only the academic knowledge but also the skills and know how working with people from all works of life. Thank you, Dr. Zhang!

I would also like to thank my thesis committee members: Professor Cinzia Cirillo, and, Professor Paul Schonfeld, for their willingness to help me accomplish my thesis research through encouragement, insightful comments, and challenging questions. I will always remember your professionalism.

I own my great thanks to my colleagues and friends in the University of Maryland. I would like to thank Xiang He and Arefeh Nasri for their knowledge and encouragement on my research. They have my gratitude for being there whenever I encountered issues and difficulties. I also want to thank Yijing Lu and other members in our group. Their friendship and encouragement make me succeed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Background Information

Land use, travel behavior and environment interact and react with each other. It is generally assumed that "rational" individuals attempt to choose their home and work location and the geographic spread of the activities in which they participate, so they can minimize their travel cost, which is often measured by travel time, distance and convenience (Joyce M Dargay & Mark Hanly, 2003). There is no doubt that land use patterns of home and geographic activities have an important impact on people's travel behavior. On the environmental front, global warming due to greenhouse gas emission and air quality nonattainment areas have led to significant effort in collecting and processing vehicle activity data, testing emission rates, and quantifying mobile emissions. Vehicle Soak-time, which is defined as the duration of time in which a vehicle's engine is not operating between trips, is an important factor in vehicle emission modeling. Exploring the distribution of vehicle soak time and how it can be modeled and predicted is of particular interest among air quality emission modelers and researchers.

There have been abundant researches involved in the environment impacts evaluation as related to travel behavior both within and outside of the United States. Population density, urban size, the proximity and frequency of public transport alternatives and distance to facilities for shopping, services, leisure, land use mix, street connectivity, the access to regional employment have been inspected on their impacts on land use (Joyce M Dargay and Mark Hanly, 2003; Lawrence Frank and Company, Inc, 2005). Census Tract and Traffic Analysis Zone (TAZ) are the mainly used geographic units of analysis in these studies because of data availability (Zegras, 2010; Boarnet andCrane, 2001) in spite of its large scale. In some special case, the scale of measurement was an entire county (Ewing et al 2003), or a metropolitan region (Sturm and Cohen 2004). However, there have been no studies on how the person daily activity space level land use factors impact travel behavior, mostly because of the data limitation.

Over the years, research has also demonstrated that vehicle soak time correlates with the vehicle start information, such as whether a vehicle's start is cold (longer vehicle soak time duration) or not (shorter vehicle soak time duration), and when it is a first start or not a first start. These variations result in large difference in vehicle start emission factors. An analysis for vehicle soak time and its distribution can better clarify the relationship between the vehicle soak time period and corresponding vehicle start characteristics of a vehicle soak period. Such data will provide more accurate inputs to vehicle emissions models.

Coupled with all the increasing interest discussed above, there is also an increasing availability of travel survey data. Of particular interest is the travel survey data collected in different geography-levels, such as the national level, regional level, and metropolitan areas, and local level. The National Household Travel Survey (NHTS) conducted in 2009 is one of the readily available data sources that provide good information on nationwide vehicle soak time data and have yet to be fully explored. A set of GPS survey data obtained by the University of Maryland provided another detailed local data to examine such issues. The University of Maryland GPS data has a recording interval of every 1 minute and covers the Greater Baltimore-Washington Metropolitan Area. When this set of GPS data is loaded into Arc GIS and exported to a shape file, they can easily be used for the person travel information analysis. Also, the geocoded DC and Biltmore metropolitan land use data can be jointed with the GPS point data, and that can help calculate land use factors at activity space level.

1.2 Research Objectives

The primary objective of the study is to characterize how measures of residential density, employment density, land use mix, and average block size where people live and travel daily influence their travel behavior, in particular the vehicle miles travelled (VMT). Land use measures are not only developed at the origin TAZ level, but also the activity space level, which included all locations people have visited in the whole day. The GPS survey data has recorded people's travel information at different days. A regression model with panel data analysis is introduced to estimate how the different level land use measurements impact the travel behavior differently, with socio-demographic and other factors are controlled during the evaluation. The panel model allows us to control for variables that can't be observed or measured. Since vehicle emission rates are heavily influenced by vehicle soak time distributions due to their impact on vehicle start emissions and evaporative emissions, a secondary goal of the project is to look at time of day, day of week, trip purpose, vehicle type, gas price and several interaction variables impact on the vehicle soak time distribution using the NHTS 2009 data. A model analyzing the start mode fraction is built with logistic regression methods. Again, time of day, trip purpose, day of week and their interactions are found to be the significant factors explaining the differences between vehicles soak periods prior to first start and those of non-first starts. Following the start mode fraction model, the generalized Gamma model on non-first start vehicle soak time durations is also established. This model enables emission modelers and analysts to predict soak time distributions based on demographic, socioeconomic and travel behavior characteristics.

1.3 Summary of Contributions

This thesis is using GPS data to look at land use impact on travel behavior; the built environment factors are measured both at origin level and activity space level. In the meantime, the NHTS data is used to estimate and model vehicle soak time distributions. Compared with previous study, this thesis has made contributions in the following aspects:

First, instead the traditional survey data, the thesis takes the GPS survey data to inspect the relationship between built environment factors and travel behavior. Activity space, one measurement method for travel behavior, is also explored; even if the model result is not as good as the model using VMT as the travel behavior measurement. Also, the built environment factors are furthermore explored at the activity space level, which is not studied before.

Second, while relatively new, this thesis estimates and models vehicle soak time distribution using 2009 NHTS survey data. New soak time distributions can be modeled through this method by using new travel survey data. Also, the gas price is analyzed to see how it effects the vehicle soak time distributions.

1.4 Outlines of Thesis

The remainder of the thesis is organized as following.

In Section 2, available literature reviews involving in land use and travel behavior, and vehicle soak time related analysis have been conducted and serves as the foundation for the basic research idea. In the following section, 2009 NHTS and Maryland GPS survey data are first introduced. Then based on the purpose of the thesis, the author does some data processing and comes up with the final dataset. Section 4 and Section 5 focus on the model specification and results analysis. Regression model with pane analysis is used to determine how different land use measurements impact on VMT. The generalized Gamma model is introduced for the non-first start vehicle soak time analysis. Section 6 concludes the thesis and discusses the model application and future research work.

Chapter 2: Literature Reviews

2.1 Land use impact on travel behavior

It is generally assumed that “rational” individuals attempt to choose their home and work location and the geographic spread of the activities in which they participate, so they can minimize their travel time or travel distance. There is no doubt that land use patterns of home and geographic activities have an important impact on people’s travel behavior. On the environmental front, global warming due to greenhouse gas emission and air quality related health issue in air quality nonattainment areas have led to significant effort in collecting and processing vehicle activity data, testing emission rates, and quantifying mobile emissions. Vehicle Soak-time, which is defined as the duration of time in which the vehicle’s engine is not operating between trips, is an important factor in vehicle emission modeling. Exploring the distribution of vehicle soak time and how it can be modeled and predicted is of particular interest among decision makers and researchers.

As early as 1996, planners are increasingly seeking to understand the relationship between land use and travel behavior with quantitative assessment. It has been always difficult to establish the relationship and obtain conclusive evidence. Boarnet, Marlon G., and Sarmiento. Sharon (1996) were attempting to formulate and estimate a model of how travel behavior is related to land use patterns near a person’s place of residence. The final model results are not as strong as originally expected, mostly because of the sample dataset's dependence on auto, high income groups, working people. Land use factors including population density, percentage of street grid within a quarter mile radius of person's residence, retail employment density, service employment density, total employment divided by total population used to measure jobs-housing balance are suggested to be further researched.

Joyce M Dargay and Mark Hanly (2003) used seven days travel survey data which is conducted in Great Britain to analyze the same issue. Land use factors, such as population density, urban size, the proximity and frequency of public transport alternatives and distance to facilities for shopping, services, leisure, etc, are explored at home location. Total distance travelled, distance travelled by car and the total number of journeys on an individual level are analyzed. Simple regression model is applied with all individual, household. Land use variable is binary, and it equals to 1 if the condition holds and equals to 0 otherwise. In summary, it has concluded that land use characteristics do play a significant role in travel demand and car use.

Mark Bradley and Keith Lawton (2005) described land use mix, density and street connectivity where people live and work influences their trip making patterns including trip chaining and mode choice for home based work trips, home based non-work trips, and mid-day trips from work. The Puget Sound Council's 1999 Household Travel Survey data provided trip, person and household level socio-demographic data for the analysis for this research. Land use measures are developed within one kilometer of the household and employment trip ends in the survey. Multivariate linear regression models for mean daily household level vehicle miles and hours traveled were developed to investigate their relationship with land use and street network variable at the one kilometer household buffer level. The household level demographics, including number of vehicles and people per household and household annual income are controlled. Regression model results clearly indicate that land use measures play an important role in both how much we travel.

In Metin Senbil, Junyi Zhang, and Akimasa Fujiwara's paper (2005), one day travel survey data is used to derive person trip information, individual attitudes on various issues from general to specific; and geocoded land use data is added. In this thesis, the authors concluded that for short-term mobility decisions, policies drawing from land use and transportation system might be effective; however their effects on long term mobility decisions, i.e., private car ownership and commute trip mode are found to be ineffective. Further analysis is needed based on more detailed land use and transportation characteristics. The bivariate binary Probit regression model, Poisson Regression, and Tobit regression are models used in the model analysis.

In 2010, Antipova (2010) explored land use, individual attributes, and travel behavior in Baton Rouge, Louisiana. Travel survey data, geocoded commuting origin and destination information, TAZ level land use information are used in this thesis. Built environment variables, such as land use type where a respondent resides, land use mixed captured by the jobs to workers ration (JWR) around the respondent's home, and the respondent's proximity to a high-performing school, are used to measure the built environment effect. A multi-level modeling approach is applied to investigate the geographical effect of a place and the role of population composition in accounting for place-to-place differentiation in commuting, including the commuting distance and commuting times.

In Nasri and Zhang (2011)'s analysis, the authors reexamined the effects of built environment factors on travel behavior, in particular VMT in five US metropolitan areas grouped into four case study areas. The authors also developed consistent models with the same model specification and datasets for all areas to enable direct comparisons. Seattle 2006 Household Activity Survey and 2005 building parcel land use data. The analysis relied on the Greater Baltimore- Washington Metropolitan Council of Government and the Baltimore metropolitan Council, 2009 NHTS add on data and 2009 land use data from Virginia Department of Transportation. Residential density, employment density, entropy, average block size, and distance to city center are used to measure the built environment impact on travel behavior. The Bayesian multilevel model can be considered as an extension of regression models that produce different coefficients by subject groups. Subjects in the same level/group are likely to be similar to each other in terms of their observable characteristics.

2.2 Vehicle Soak time

Vehicle soak time has been explored by a number of researchers. These studies share many similarities. They used travel survey data, such as 1995 National Personal Travel Survey (NPTS) data and Dallas metropolitan area household travel survey data. They all modeled vehicle start types, simply using the time-of-day and trip purpose variables. Some later researchers advanced this model specification considerably by incorporating some zonal attributes. All the studies are related to EPA's emissions models, most notably, the MOBILE models.

Motivated by EPA's factor models (EPA 1994, 2003), many earliest analyses focused on start mode fractions and model vehicle soak time distributions as the input to emission forecasting. Venigalla and Pickrell (2002) have modeled the aggregate proportions of cold and hot start using the data from the National Personal Transportation Survey (NPTS). The variation in start mode fractions has been discovered by this thesis as closely correlated to a driver's trip purpose and time of day attributes. Their approach can be used to provide the start fractions needed by MOBILE5. On the basis of a detailed statistical analysis, a grouping scheme is devised to consolidate vehicle soak distribution inputs by time period. The grouping scheme will enhance the utility of survey data in deriving the vehicle soak distributions and reduce the effort in providing vehicle soak distribution inputs to MOBILE6.

Nair et al. (2002) also studied the data from household travel survey conducted in the Dallas area and other supplemented data sources. Log-linear regression and logistic regression have been used in this study. Their approach used similar independent variables compared to Venigalla et al. (2002), but focused on modeling the disaggregate vehicle soak-time distributions, which are important input of the MOBILE6.

Gao and Johnson (2009)'s paper reviewed statistical analysis methods available in practice or still in research, whichever are relevant to analyzing vehicle soak time data. They provided valuable comments on pros and cons and theoretical justification of these methods. The thesis is finalized by a general guideline for the analysis of vehicle soak time data. Finally, a subset of the statistical methods discussed is used to analyze the US EPA’S 3-city data.

Venigalla and Miller (2002) classified trip start as cold start and hot start. Emission modeling programs use these start modes as direct or indirect inputs to procedures or models that would be used to determine the portion of vehicle miles traveled in transient and stabilized operating modes. They also refer that trip purpose is the most important explanatory variable for variance in cold starts, followed by the temporal variables such as the time of day at which the trip is made. The start mode fractions are useful for a variety of mobile source emission modeling.

MOBILE 6 is a computer program that estimates some emission factors for gasoline and diesel highway motor vehicles, and other specialized vehicles. The input parameters for the MOBILE6 include calendar year, month, and vehicles characteristics. Engine starts vehicle soak time distribution by hour, which is our focusing point in the research are key inputs for MOBIL 5. Outputs include emissions from hot vehicle soak conditions, running vehicle soak and so on. MOVES2010a, released in August 2010, is EPA’s state-of-the-art tool for estimating emissions from highway vehicles, which can also be used to analyze emission rate by inputting vehicle soak time duration related variables. The MOVES also includes spreadsheet tools that help with the transition from MOBILE6.2 to MOVES.

2.3 Summary of the literatures

Above all, topic about the relationship between built environment factors and travel behavior has been explored by significant number of researches in recent years. Specifically, density, land use mix, accessibility, average block size, distance to CBD, all kinds of service areas, are the mainly considered built environment factors. These factors are calculated based on the travelers' home or work locations. National travel survey data is the mostly used data source. The land use analysis focus on the census tract level or TAZ level, in which the data is more easily obtained. Moreover, some analysis is using the county level or metropolitan level geography as the analysis zone.

In my thesis, both the national travel survey data and regional survey data are used. There are important differences between this research and what other people have done. First, the author not only analyzed trip origin land use impact on travel behavior, but also evaluated the average land use variables at activity space level factors. That is because the GPS survey data can record people travel information at one minute interval, that can help use obtain the exact people travel information at each day. My thesis focuses on the GPS data processing to obtain and calculate a person's origin and activity space level built environment factors. Regression with panel data analysis is used as the modeling procedure. Both time-fixed effect and person-fixed effect are discussed in the modeling part.

For the vehicle soak time analysis, some researches have tried to find the factors that may impact the vehicle soak time distributions. My thesis has analyzed the disaggregate vehicle soak time distributions, and the start pattern as a function of a series of interaction variables in addition to time of day and purpose. Furthermore, vehicle soak-time distributions for first starts and non-first starts of the day are explicitly distinguished because the distributions for these two types of starts are likely to be very different.

Specifically, in the vehicle soak time analysis, the thesis firstly focuses on 1) difference of first and non-first start vehicle soak time distribution, and 2) vehicle soak time distribution considering the time of day. For the second part, relationship between soak time duration and some corresponding variables has also been analyzed. Vehicle soak time distributions have also been evaluated by using the regression model to determine whether a trip is a first start trip or non-first start trip.

Chapter 3: Study Area and Data Processing

3.1 Maryland GPS data and NHTS data introduction

Our GPS survey data is driver level survey conducted by the University of Maryland. The GPS device can record travel data in one minute interval. It also provides very specific travel data for each person at each day. Based on the dataset, a person's travelled location and even frequencies of locations where people have visited can be obtained. Also, a trip is divided if its velocity is equal to 0 for two minutes. The distance is calculated by each two points based on the coordinate system.

The National Household Travel Survey with the weighted data is household-level survey covering all the United States. It provides data on personal travel behavior, trends in travel over time, trip generation rates, national data to use as a benchmark in reviewing local data, and data for various other planning and modeling applications. EIA also has the monthly gas prices file for all states from 1990 to 2011.

For our research purpose, the GPS survey data is used to compute the travel distance for each person at each day. The NHTS survey data would be the primary source to compute vehicle soak time. In addition, by spatial connection, the land use data can be joined with the GPS data. This can help us conduct analysis the land use impact on travel behavior, in particular the travel distance. Meanwhile, by keeping a unique household ID, vehicle ID and person ID for each vehicle soak period, vehicle soak time file and vehicle file and person file can be joined. This has allowed us to conduct analysis on the impact of household, vehicle and person characteristics on vehicle soak time duration and distributions.

3.2 Maryland GPS data Processing

Before coming up with the cleaning dataset the project needs, special processing and compilation is needed. The steps listed below illustrated the process adopted.

Step I: Load the raw GPS data into Arc GIS to get Geo-coded shape file.

The most important thing in this procedure is to project the data correctly for its geospatial location. The author used the projected coordinate system of the land use dataset as the base coordinate system. Because of the large sample size, the data is split into 16 files.

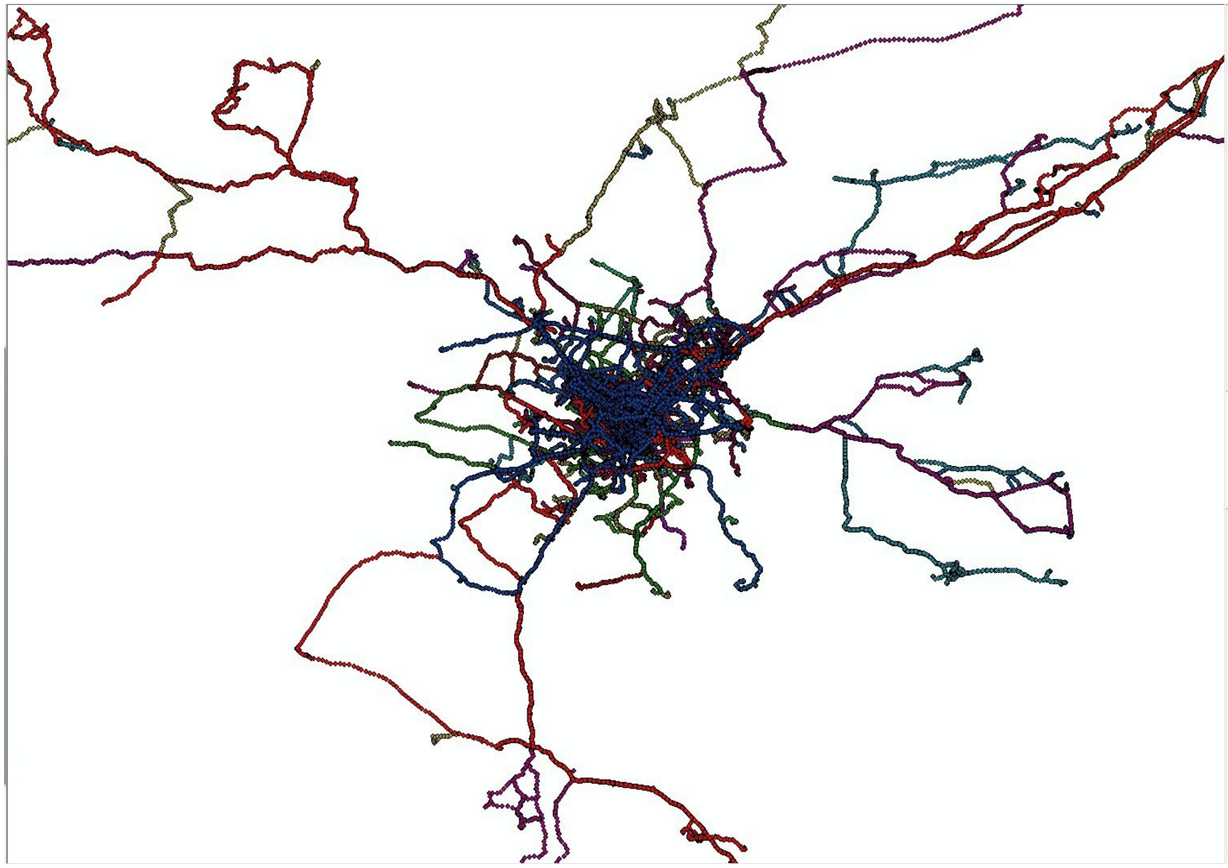


Figure 3-1: All GPS point data

Step II: Spatially join each GPS data file and the DC and Baltimore metropolitan land use file

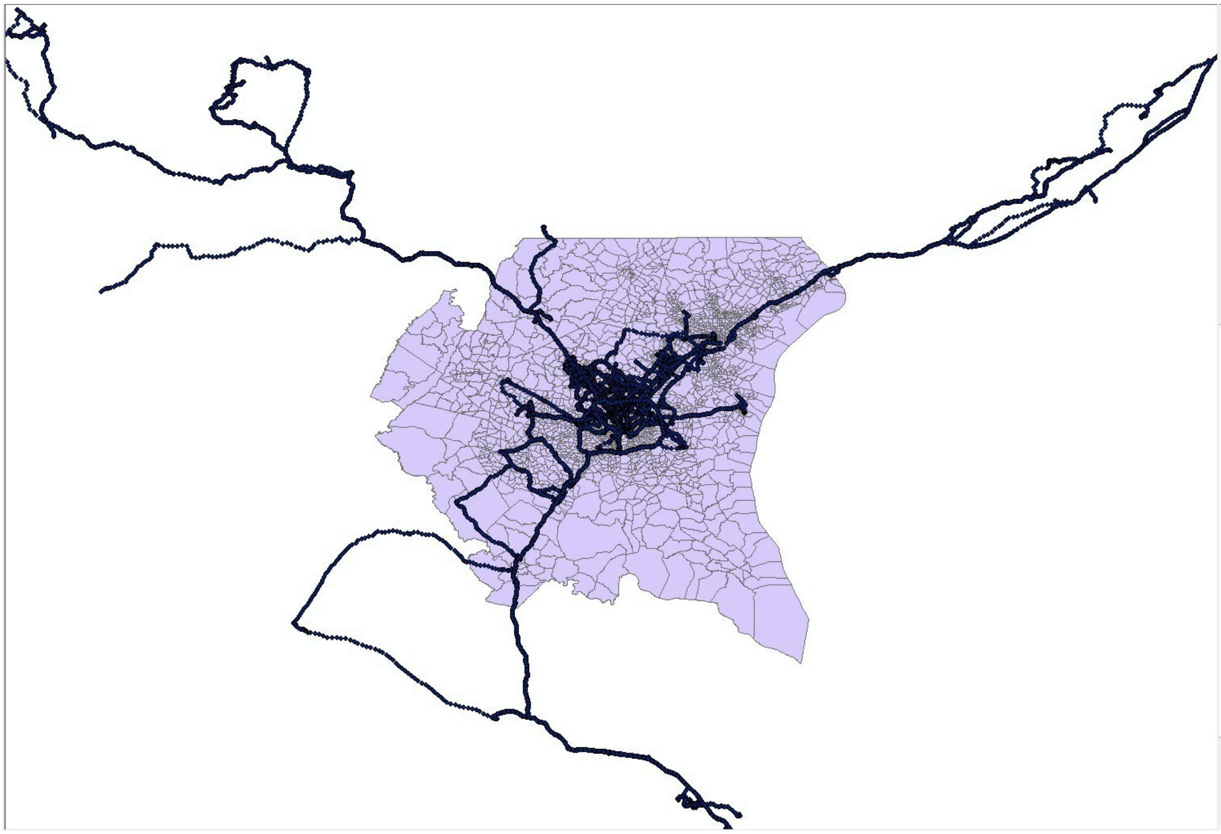


Figure 3-2: GPS data points spatially joined with BMC and DC land use data

Step III: Identify and exclude trips which are outside analysis area

Since the land use dataset only included the Baltimore and DC metropolitan area, travel trips outside the research area are excluded because of the land use data limitation. Specifically, if one person has travelled outside the research area, his or her trip for the whole day is removed. That is because the research is based on person's daily travel information.

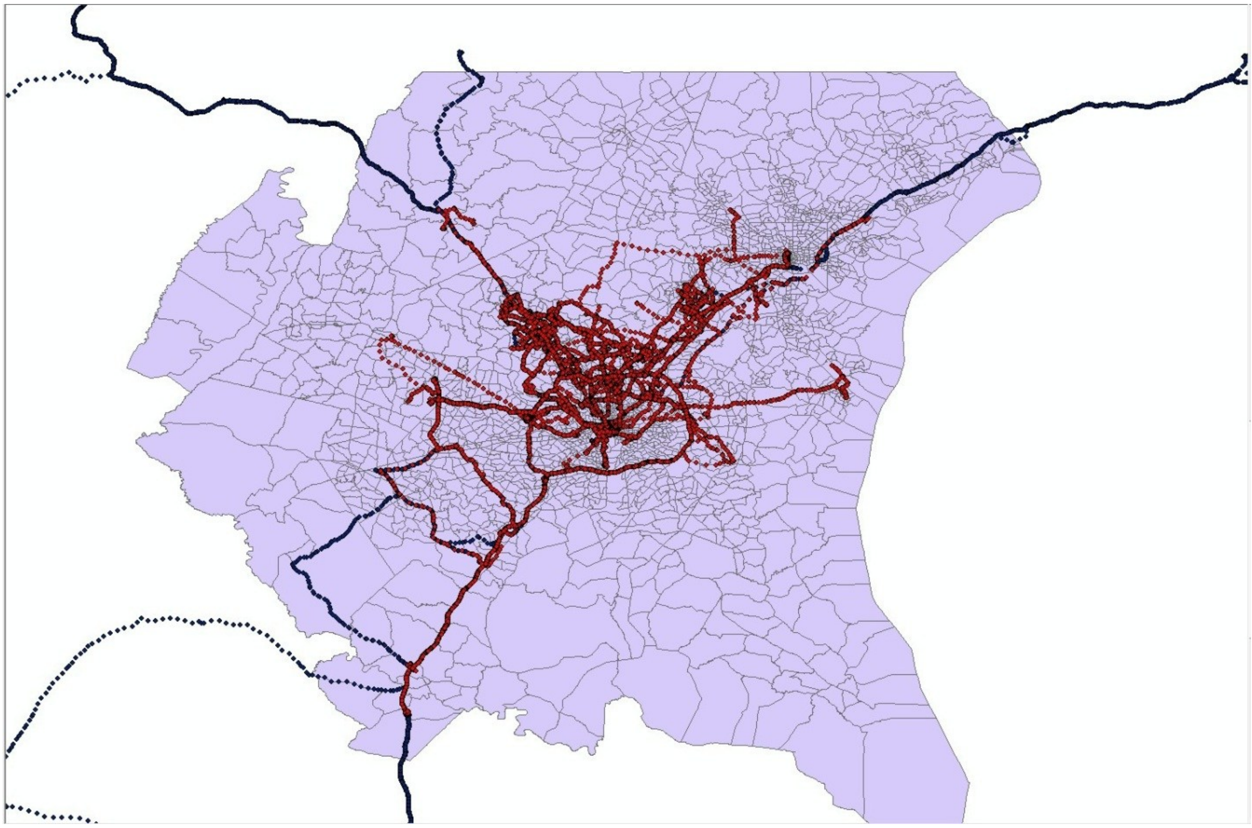


Figure 3-3: Travel within BMC and DC MPO area

Step IV: Calculate density, land use mix and average block size

First, residential and employment densities are calculated based on the number of residents or employments within each TAZ area. The two density measures were calculated by using the following formula:

$$Y = X / \text{TAZ area}$$

Where, depending on the density measure being calculated:

Y= residential density or employment density

X= number of residents or employments per each TAZ area

Second, the land use mix was calculated based on a measure of entropy between residential and employment land uses. The equation results in values between 0 and 1. The closer to 1 the values is the more evenly distributed the building floor areas are between uses. A value of 0 indicates a single use. For those who have 0 residents or employments at this TAZ area, the approximation is any value equal to 1 is treated as 0.000001. The entropy is calculated based on the following formula.

$$\text{Entropy} = -\sum_j \frac{P_j * \ln(P_j)}{ln(J)}$$

Pj = the proportion of land use in the jth land use category

J = the number of different land use type classes in the area

Finally, the average block size is calculated at TAZ level.

Step V: Based on the dataset, the daily travel origin land use variables are obtained and the average activity space level land use variables are calculated.

With the help of Arc GIS software, the TAZs, which contain all GPS points, are picked up. Then the author calculated land use factors for each TAZ. Finally, the average land use factors are calculated based on the previous selected TAZs. Since the visited locations have the different frequencies, time travelled at each TAZ for one person at each day is used as the weight factors for each TAZ level built environment factors.

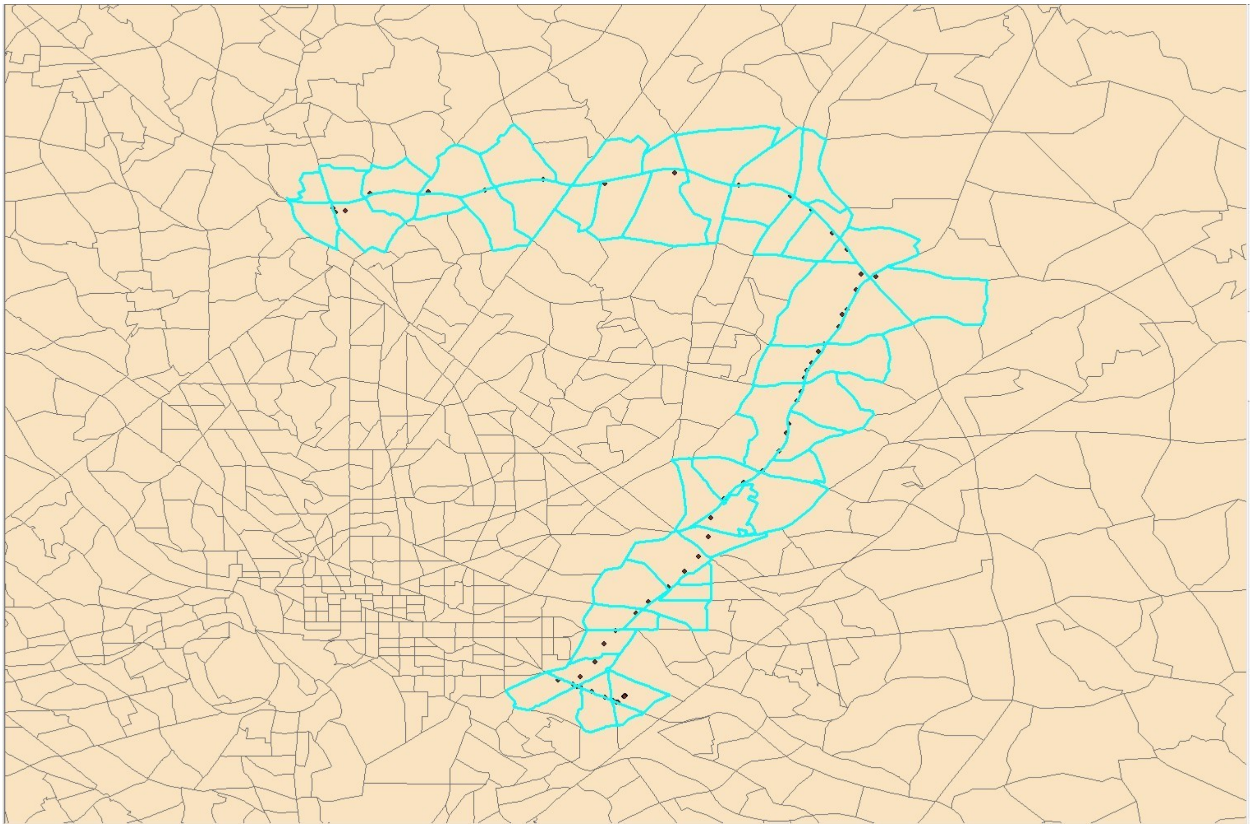


Figure 3-4: One day travelled TAZ area

Distance is calculated based on two GPS points. Trip length can be calculated by summing all two points' distances. The VMT used in my thesis is calculated by adding all trips of the whole day for one person together.

Trip level land use measurement can also be explored in the future. Destination choice, route choice, and trip generation rate are good representatives for the travel behavior.

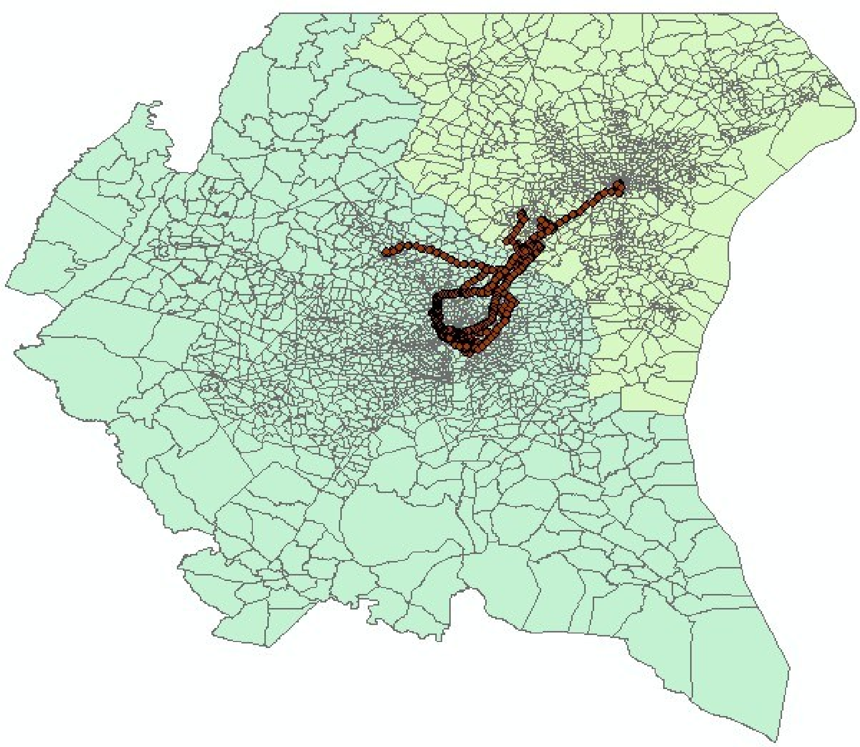
Step VI: Join the socio-demographic factors for each person.

Finally, each person has each day's travel information and corresponding TAZ level built environment variables in the dataset.

Step VII: exploring different activity space level land use factors measurement methods

Convex Hull in minimal boundary geometry is used for the construction of activity space. The size or area of the activity space represents the measurement of travel behavior. This means that the polygon created is the smallest possible area shape that is not a concave segment (an internal angle greater than 180 degrees). This was selected over more simplified methodologies due to the tight fit around the points it creates without the error associated with a concave segment.

Kernel density is also an activity space calculation method that has been used by others. That kind of land use measurement method at activity space level can be explored in the future.



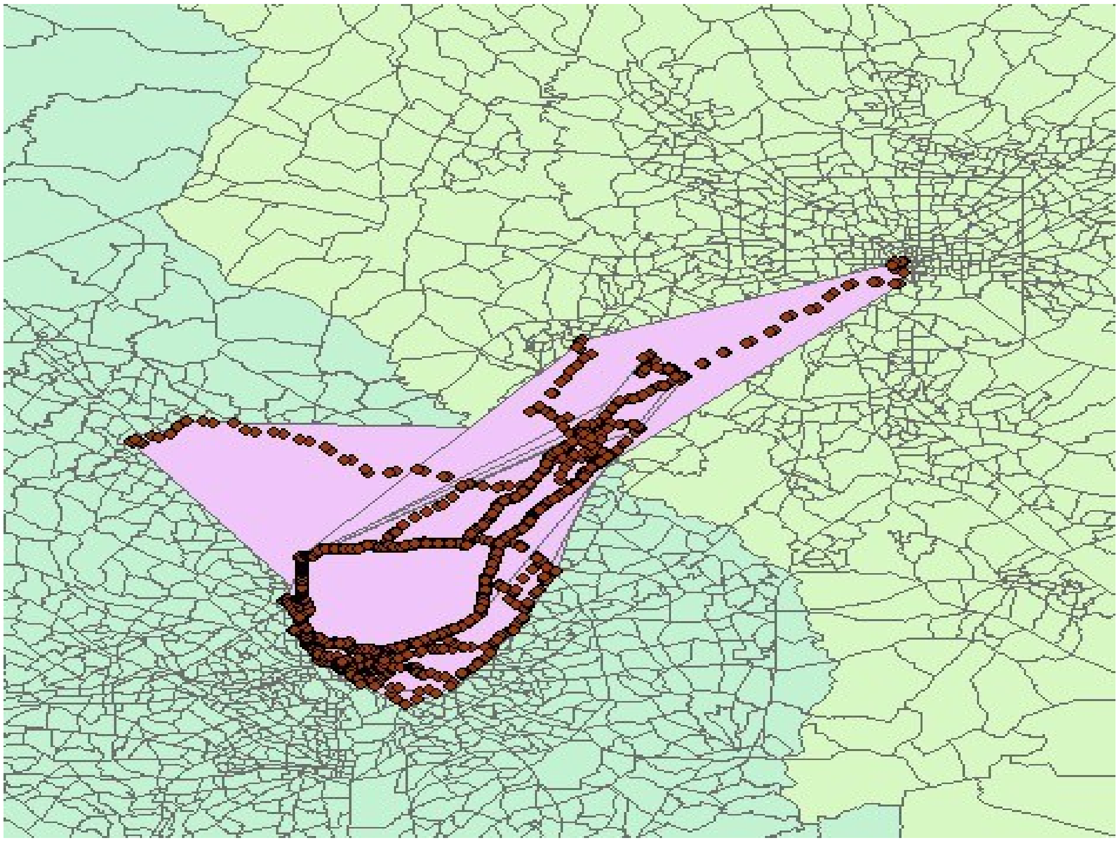


Figure 3-5: activity space calculation procedure

3.3 Vehicle Soak time data processing

For the NHTS data, there are also some issues relating to the travel day file, which would cause some problems when computing vehicle soak time from it. One of the major issues is the overlapping of trip duration. As the travel day file is recorded trip by trip, there would be some cases that more than one respondent take the same trip together in the same vehicle. And in most of these cases, the respondents would not report exactly the same start and end time of that trip due to slight differences in personal perceptions. In this situation, we have to merge that trip records into one record.

Generally, the data processing was conducted following the steps listed below.

Step I: Delete the trips with TRPHHVEH<>1

TRPHHVEH represents whether household vehicle was used on that trip, and the value range code of this item is -1(Appropriate Skip), -7(Refused), -8(Don't Know), -9(Not Ascertained), 1(Yes) and 2(No). To analyze the vehicle soak time distributions, we need to know all the trips taken by a particular vehicle during a whole day. Obviously, the household travel survey could provide the information of all trips made by the household owned vehicles, while information of only part of the trips information made by public transportation or vehicles from non-respondent households could be gotten. Thus we only keep the information of trips made by vehicles from respondent households for our analysis.

Step II: Identify trips with VEHID<0 and delete all the trips made by the corresponding household.

The remaining records are all trips taken by respondent household owned vehicles, and VEHID<0 means that we could not identify which vehicle owned by the household is used for that trip. Thus that particular trip may influence the vehicle soak time distribution of any one of the household owned vehicles. In this case, we delete all the trips made by that household from the dataset.

Step III: Identify trips with negative start or end time and delete all the trips made by the corresponding vehicle.

When the respondents are not sure about the start or end time of a particular trip, a negative value would be entered into the corresponding cell. This kind of trip would either influence the vehicle soak period preceding that trip or following that trip. Thus we delete all the trips made by that vehicle.

Step IV: Process overlapping on trip time and compute vehicle soak time.

Basically, there would be two types of cases that two trips taken by the same vehicle overlap with each other. The two cases are depicted in Figure 3-6.

A part of each of the trip overlaps in case 1, and one the trips covers the whole period of the other trip in case 2. In both cases, the two trips are merged into one trip with the earlier start time and later end time of the two trips as the start and end time of the newly merged trip.

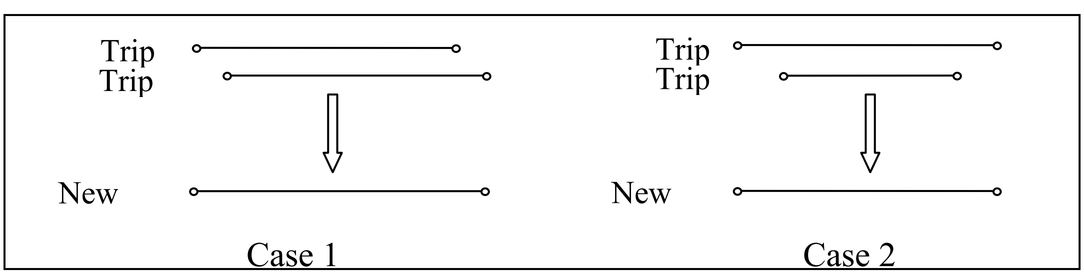


Figure 3-6: Overlapping of Trips

When the merging of trips is completed, the start and end time of each vehicle soak period could be easily gotten.

As the NHTS only collects data in a 24-hour period, determining the vehicle soak time before the first trip of a day made by each vehicle would be very difficult. In our analysis, we assume that characteristics of the previous day's travel activity are identical to those of the observed travel day. Then the vehicle soak period before the first trip could extend to the same time on the previous day as that the last trip ends on the observed day.

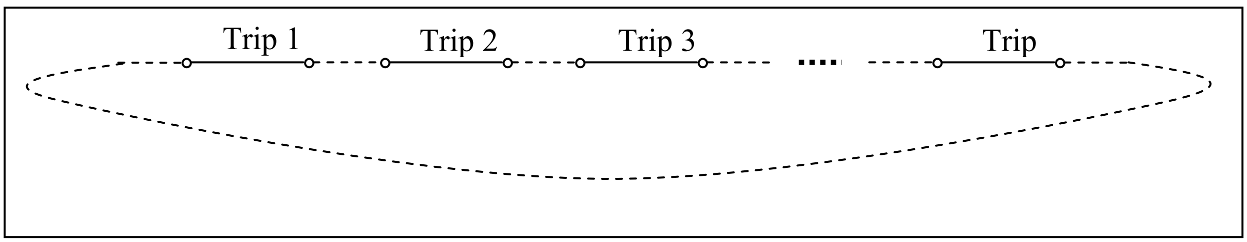


Figure 3-7: Vehicle Soak Time along a Whole Day

Step V: Delete vehicles which have only one vehicle soak time record.

There would be some cases that only one trip was made by a vehicle. Under our assumption that travel pattern is the same on the previous day as that on the observed day, there would be only one vehicle soak time record. However, we know that it's not normal a vehicle only makes one trip during a day, the assumption is not reasonable in this situation. To make our statistics of vehicle soak time more realistic, we'd rather delete all the vehicles with only one vehicle soak time record.

Step VI: join the derived vehicle soak time records with household, vehicle, person files and EIA gas prices file.

For the purpose of exploring vehicle soak time distribution by different characteristics, we need to join the derived vehicle soak time dataset with household, vehicle, person files and EIA gas prices file to get the corresponding household, vehicle, person, and gas prices attributes of each vehicle soak time records.

Chapter 4: Regression model with panel data analysis for the land use impact on travel behavior

4.1 Introduction

For this study, the dataset is from the University of Maryland GPS survey data which was conducted in the 2011. The sample size after cleaning for my analysis consists of 164 individuals, which produced 7118 daily travel trips. As indicated in Table 4-1, it is found that male respondents are a 12% more than female respondents. Middle age respondents explained the most sample, and most people have four year college education level or higher than that. Only 9.76% respondents have the yearly income lower than 50000.

Table 4-1: Summary of Selected Demographic Variables

	Frequency	% Share
Gender		
Male	95	57.93%
Female	69	42.07%
Age		
age between 17 and 34	28	17.07%
age between 35 and 64	134	81.71%
age higher than 65	2	1.22%
Education		
High School Graduate	2	1.22%
bachelor master and phd	140	85.37%
Associate and College	22	13.41%
Income		
income less than 50000	16	9.76%
income between 50000 and 100000	50	30.49%
income between 100000 and 150000	52	31.71%

income higher than 150000	46	28.05%

Table 4-2 presents the descriptive statics for the built environment variables in the trip origin and average activity space level. In general, the built environment factors in this two measured methods are similar with the mean value. Only for the employment density, the value at the activity space level is higher. These descriptive statistics are encouraging because cases with different measurement of built environment features are ideal for this study.

Table 4-2: Summary of land use variables

Origin built environment factors	**Mean**	**Median**	**S.D.**
Origin TAZ Population density	6.8871	4.0813	5.3352
Origin TAZ Employment density	5.6566	3.7112	14.7846
Origin TAZ land use mixed entropy	0.3226	0.309	0.2616
Origin TAZ Average Block Size	0.30899	0.245	0.9403
Average built environment factors	**Mean**	**Median**	**S.D.**
Average TAZ Population Density	7.2252	6.3093	4.0695
Average TAZ Employment Density	9.9439	5.6336	13.4888
Average TAZ Entropy	0.5001	0.3482	0.6815
Average TAZ Block Size	0.6233	0.4198	0.5318

The table 4-3 summarized the results about the built environment factors impact on VMT.

Table 4-3: Hypothesized built environment impact

Measure	**Definition**	**Hypothesized impact on VMT**
Residential density	Population/Area size	Negative

Employment density	Employment/Area size	Negative
Land use mix (Entropy)	Mixture of residential and employment	Negative
Average Block size	Average block size within TAZ	Positive
"Negative" means higher residential density leads to lower VMT per person, which is desirable.		
"Positive" means larger block sizes leads to higher VMT per person, which is desirable.		

4.2 Hypothesis test for the fixed and random effects model

The Maryland GPS survey data is a longitudinal dataset where the personal behavior is observed across time. Panel data allows us to control for variables that can't be observed or measured like cultural factors or difference in preference across people; or variables that change over time but not across person.

Usually, when we analyze panel data, two techniques are used. One is the fixed effects, and the other is the random effects analysis. You may apply regression model with entity fixed effects when you want to control for omitted variables that differ among panels but are constant over time. Meanwhile, for unobserved effects that vary across time rather than across person that impact the travel behavior, you can run regression model with time fixed effect to control for such unobserved variable that may vary by time. The random effects model assumes that an independent variable is random and generally used if the level of the independent variable is thought to be a small subset of the possible values which one wishes to generalize.

To decide between fixed or random effects, we can run a Hausman test. It basically tests whether the unique errors (u_i) are correlated with the repressors; the null hypothesis is they are not. Run a fixed effects model and save the estimates, then run a random model and save the estimates, then perform the test.

The following code is used in the STATA to test the hypothesis.

```
xtset  id finaldate
       panel variable:  id (unbalanced)
        time variable:  finaldate, 40865 to 40964, but with gaps
                delta:  1 unit

xtreg distance popden empden entropy blocksize  avgweightpop avgweightemp
avgweightentropy  avgweightbs, fe

estimates store fixed

xtreg distance popden empden entropy blocksize  avgweightpop avgweightemp
avgweightentropy  avgweightbs, re

estimates store random

Hausman fixed random
```

Table 4-4: Hypothesis for fixed and random effect

	(b) fixed	(B) random	(b-B) Difference	Sqrt (diag (V_b-V_B) S.E.
origin resi den	-0.0017	-0.0013	-0.0003	0.0002
origin emp den	0.0005	0.0005	1.52E-05	0.00003
origin entropy	0.0377	0.0357	0.0019	0.0054
origin block size	0.0102	0.0093	0.0008	0.0010
average resi den	-0.0155	-0.0159	0.0004	0.0003
aver emp den	-0.0025	-0.0025	-0.00003	0.00008
aver entropy	-0.1398	-0.1421	0.0022	0.0016
aver block size	0.1197	0.1081	0.0116	0.0095
b = consistent under Ho and Ha; obtained from xtreg				
B = inconsistent under Ha, efficient under Ho; obtained from xtreg				
Test: Ho:	difference in coefficients not systematic			
	chi2(8) = (b-B)'[(V_b-V_B)^(-1)](b-B)			
	37.29			
	Prob>chi2 = 0.0000			

Since the Prob>chi2 is equal to 0.0000, and it is smaller than 0.05. The null hypothesis is rejected and the fixed effects model is statistical valid.

After the fixed effects model is defined, entity fixed effect and time fixed effect model can be chosen for the fixed model analysis. In order to make the model more accurate, we need to test the effect of time and entity. The test used performs Wald tests of simple and composite linear hypothesis about the parameters of the most recently fit model.

For the time fixed effect test, the code used is listed below.

testparm monday tuesday wedn thurs fri saturaday

(1) monday = 0

(2) tuesday = 0

(3) wedn = 0

(4) thurs = 0

(5) fri = 0

(6) saturaday = 0

F(6, 7093) = 13.57

Prob > F = 0.0000

The following is the entity fixed test.

Testparm person1 person2 person3 person4……

(1) Person 1= 0

(2) Person 2 = 0

(3) Person 3 = 0

(4) Person 4 = 0

(5) Person 5= 0

(6) Person 6 = 0

.

.

.

Two test results indicate that the null hypothesis is rejected. Time and person fixed effect variables are needed to make the model results reasonable.

We introduced different level built environment factors impact on travel behavior. Both the original built environment factors and the average land use factors are introduced in the modeling process. Socio-demographic variables, such as age, income, education level, gender and so on, are also included in the model. Dummy variables both for time and for person are added into the model as the inspection variables.

4.3 Model Specification

A multiple regression model with time and entity fixed effects panel data analysis will be used as my model. The final model specification is listed below:

The fixed effects model is seen by using binary variables. The equation for the fixed effects model becomes:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \underbrace{\boxed{\gamma_2 E_2 + \ldots + \gamma_n E_n}}_{\text{Person dummy variables}} + \underbrace{\boxed{\delta_2 T_2 + \ldots + \delta_t T_t}}_{\text{Time dummy variables}} + u_{it}$$

Where

Y_{it} is the dependent variable (DV) where I = entity and t = time.

$X_{k,it}$ represents independent variables (IV),

β_k is the coefficient for the IVs,

u_{it} is the error term

En is the entity n. Since they are binary (dummies) you have n-1 entities included in the model

$\gamma 2$ is the coefficient for the binary repressors (entities) .

Tt is time as binary variable (dummy), so we have t-1 time periods.

δ t is the coefficient for the binary time repressors .

We could also add time effects to the entity effects model to have a time and entity fixed effects regression model.

Control for time effects whenever unexpected variation or special events my affect the outcome variable.

4.4 Model estimation results and discussion

Regression model results with panel analysis are shown in Table 4-5.

Table 4-5: Regression model with time and entity fixed effects panel data analysis

Number of obs	7118	
Prob > F	0	
R-squared	0.3746	
Adj R-squared	0.3587	
Travel Distance	**Coef.**	**P>t**
Built environment factors		
origin residen den	-0.0015	0.059
origin emp den	0.0004	0.129

origin entropy	0.0322	0.157
origin block size	0.0112	0.06
average residen den	-0.0154	0
aver emp den	-0.0024	0
aver entropy	-0.1407	0
aver block size	0.0840	0.062
Monday as the base time variable		
Tuesday	0.0329	0.017
Wednesday	0.0488	0
thurs	0.0705	0
fri	0.0564	0
saturaday	-0.0206	0.163
sunday	-0.0648	0
Age factors		
age	0.0405	0.017
age*age	-0.0003	0.029
high school education as the base variable		
college and associate	-1.0532	0
bachlor/master/phd	-0.8363	0
Income less than 50000 as the base variable		
income between 5000 and 10000	0.1936	0
income between 10000 and 15000	0.9187	0.021
income higher than 15000	0.4079	0
other factors		
male	0.3023	0
Home year	-0.013	0
Hh num	0.1654	0
cons	2.4344	0

From the results, we can see that the current socio-demographic variables have significant influences on person daily VMT. For the age analysis, numerical variables are analyzed. As age grows, people travel more because of the family and work related event needs. However, the effect of age is non-linear, indicating that senior people will eventually drive less after they reach certain ages.

The education variables are introduced with high school and other education level as the base variables. Since the sample size is not large enough, with 2 high school educated people participating in the GPS survey, the base variables seems biased. If excluding the high education level sample, it seems people with four years bachelor, master, or even higher education still have a longer travel distance than those people who own the two years college or associate degree. It is most likely that high education people are tend to be engaged in spatially more dispersed business activities and social and recreational activities.

The lowest income group with income less than $ 25000 per year is introduced as the base variable. Drivers with high income travel more; but people with income between 100000 and 150000 produce the most VMT.

In terms of gender effects, male drivers travel more than female drivers. Household with more members drive more mostly because they are involving more activities. Households which have long residence time produce less VMT because they are familiar with the surrounding environment. They usually have fixed travel routes which produce less travel distance.

Monday is used as the base variable when the time dummy variables are introduced.

On weekends, drivers usually have less VMT than on weekdays. People usually have a higher travel distance on Thursday and Friday than other days of the week.

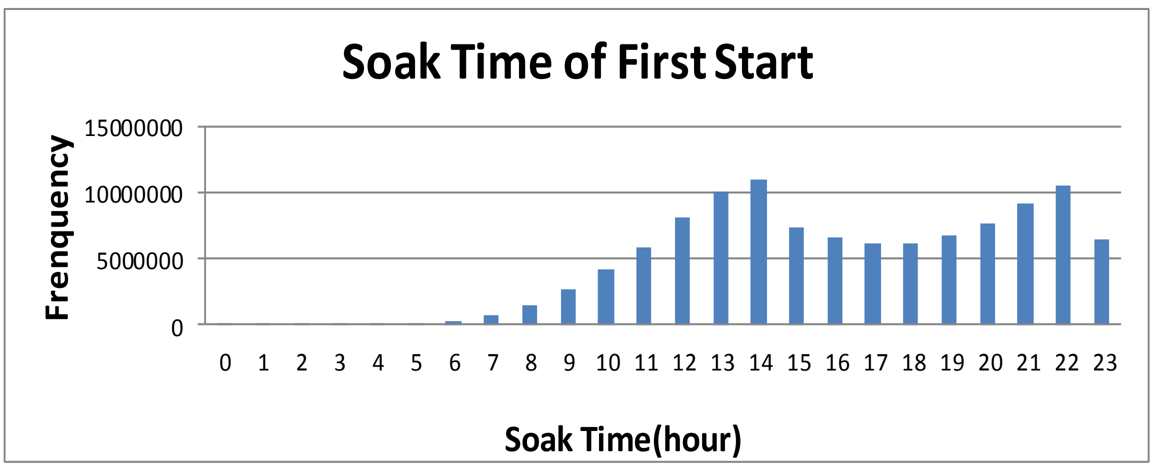
Most built environment factors have significant impacts on person daily travel distance. In my model, residential density has a significantly negative impact on VMT. Employment density is significantly correlated with VMT at the activity space level measurement. However, in the origin, the employment density has a positive impact on VMT, even though the impact is not significant. Entropy or mixed land use development have a significant negative impact on VMT at person daily travel activity space level, but not significant at the origin. Average block size has the positive effect, because a smaller block size indicates better street connectivity and workability. Comparing the origin and average residential density impact on VMT, we can see that the activity space level residential density has a larger impact on travel behavior. The above phenomena indicate that people's travel choice is more sensitive to the surrounding environment than the travelling origin land use factors. The same situation happened in the average block size impact on person daily VMT.

Chapter 5: Vehicle Soak time distribution and estimation

5.1 Vehicle Soak time distribution

5.1.1 Vehicle Soak time Distribution for first start VS. non-first start

A vehicle soak period preceding its first start should be relatively long as compared to a vehicle soak period preceding a non-first starts. Meanwhile, as the distribution of first starts and non-first starts by time of day would be significantly different, the distribution of vehicle soak times for first and non-first starts would also differ a lot along the time of day.



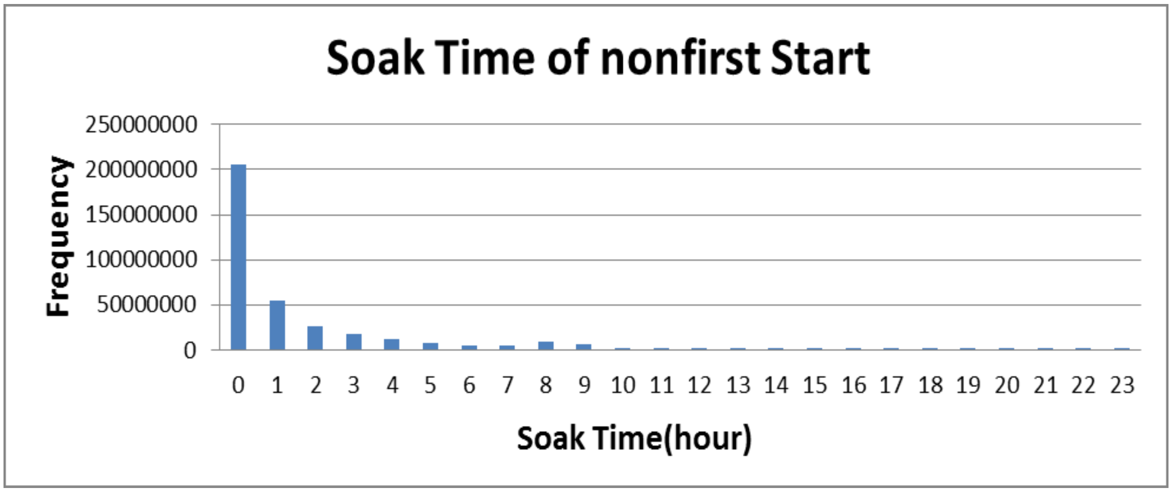
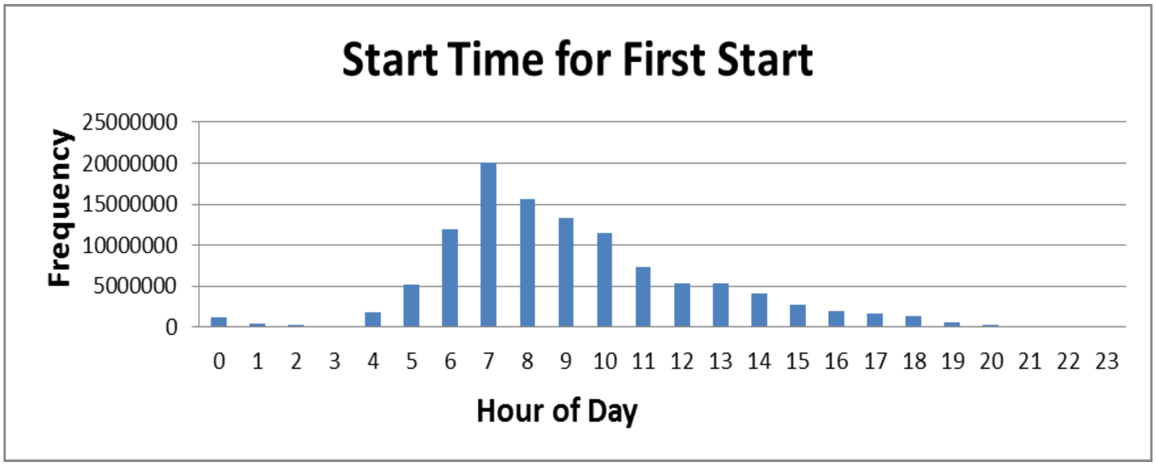


Figure 5-1: Distribution of Duration of Soak Times

Figure 5-1 indicates the distributions of duration of vehicle soak time for first starts and non-first starts. Based on EPA's model MOBILE6.2 model, a cold start is defined as a successful vehicle start following a vehicle soak time of 12 hours or more. The distribution indicates that 75.6% of the first starts are cold starts and only 0.56% of the non-first starts are cold starts.

The starting time of first starts follows a normal-like distribution at the AM peak period, while the starting time of non-first starts follows a normal-like distribution along the whole day excepting night time from 0am to 5am.The significant differences in the distribution of vehicle soak time for first starts and non-first starts suggest that we should consider them separately.

The time of day would influence the number of trips, which could be seen from Figure 5-2. In addition, it would influence the average vehicle soak time and the standard deviation of vehicle soak time as well.



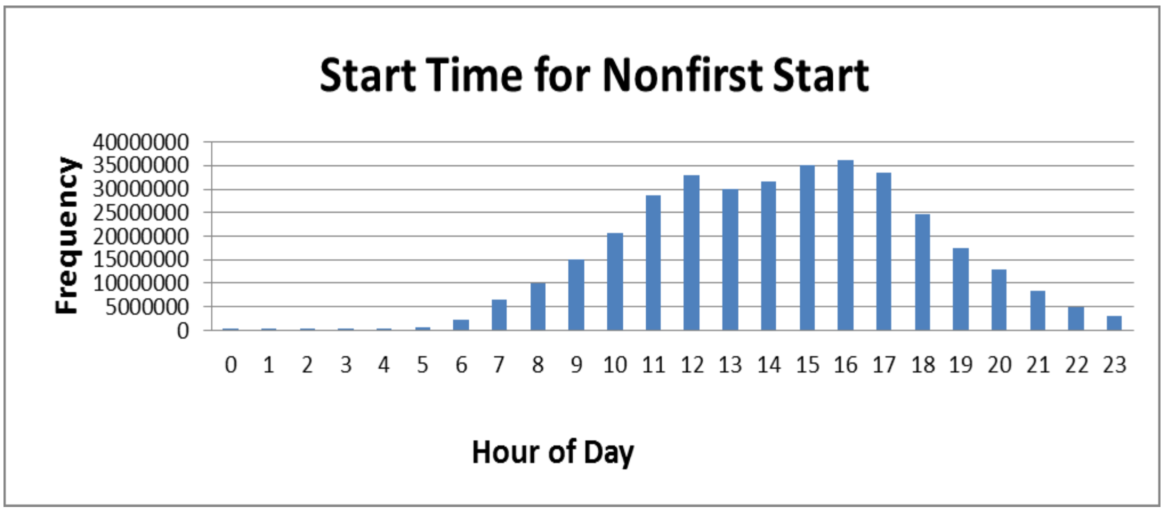


Figure 5-2: Distribution of Vehicle Start Times

The average vehicle soak time for the non-first starts during PM peak period is relatively high, which represents the activities of leaving work place for. After lunch is done, there is a long time period where people stay in the office or places in the afternoon. Vehicles soak time durations are much longer in the 5 pm until people leave after they finish the day's work. The long vehicle soak time at late night represents the activities of going back home from work or pleasure. People do not go out often in late night.

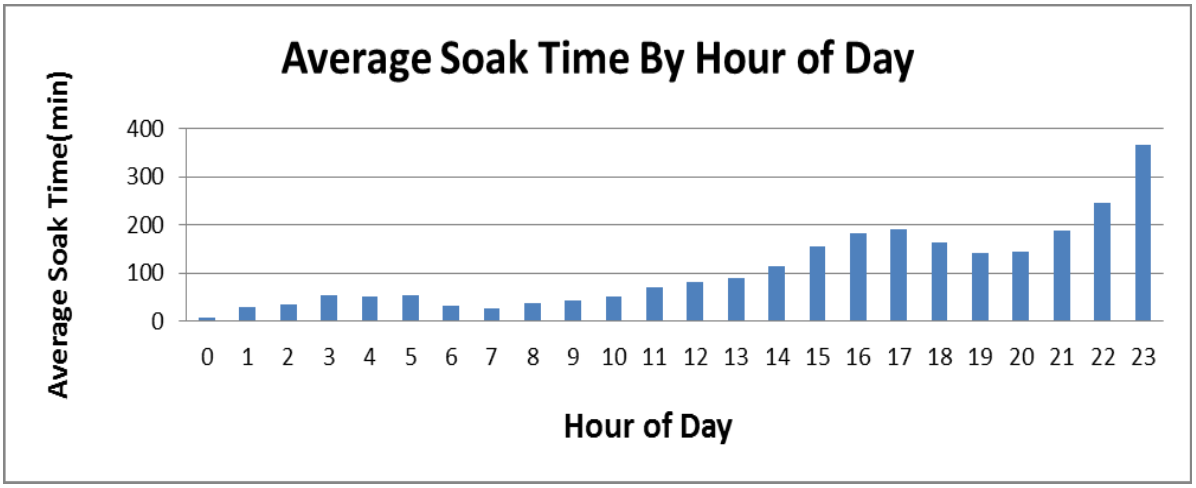


Figure 5-3: Distribution of Average Soak Time by Time of Day

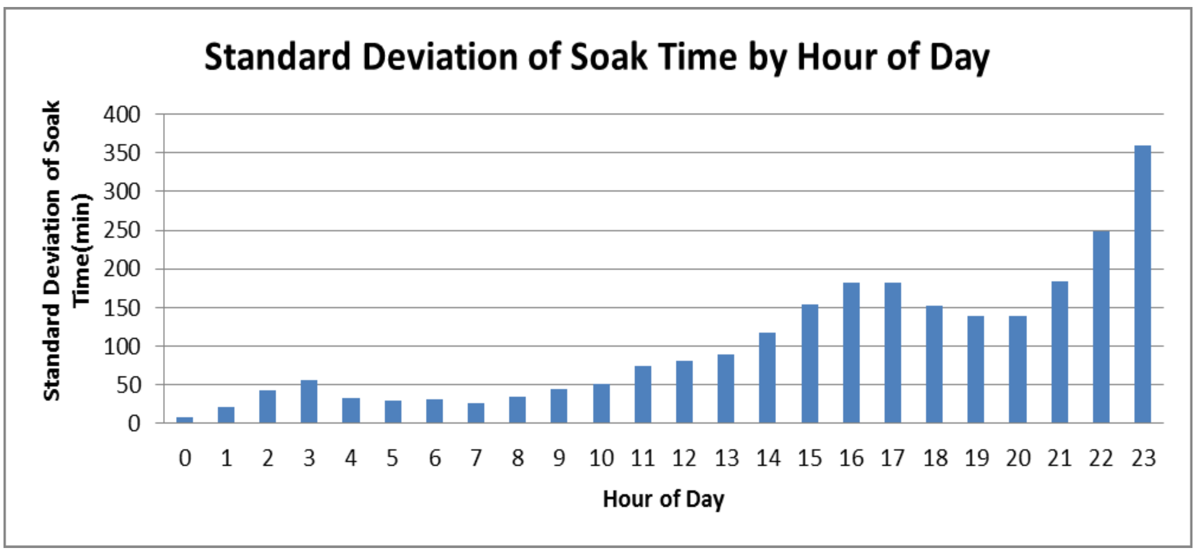


Figure 5-4: Distribution of Standard Deviation of Soak Time by Time of Day

A potential application of the analysis of vehicle soak time distributions is to derive the inputs for some existing mobile source emissions models. By comparing Figure 5-3 and Figure 5-4, we could find that the ratio of average vehicle soak time to the standard deviation of vehicle soak time along hour of day. The ratio is very close to 1 and for large number of time periods, the ration is lower than 1. This phenomenon means that the vehicle soak time distributions are not stable along time of day. To improve the quality of prediction, two methods could be implemented are implemented to complement each other simultaneously.

The first method is to incorporate other factors besides time of day when predicting vehicle soak time distribution.

Although average vehicle soak time changes significantly along time of day, it could not be directly predicted using time series models. Some other factors such as trip purpose, vehicle type, and household income level et al. would also influence the vehicle soak time. These factors may contribute a lot in the deviations.

The second method is to divide the vehicle soak time into time bins.

Instead of directly setting the duration of vehicle soak time as the dependent variable, we could divide the vehicle soak time into several bins based on the input requirements of existing mobile source emissions models. The MOBILE6 interval classification of vehicle soak time is defined in Table 5-1.

Table5-1: MOBILE 6 Soak Time Input Format

Vehicle Soak Interval (N)	Range of Vehicle Soak Time
1 to 30	Greater than N-1 and Less or equal to N minutes
31 to 45	Greater than 2(N-1) and Less or equal to 2N minutes
46 to 67	Greater than 30N-1290 and Less or equal to 30N-1260 minutes
68	Greater than 720 minutes

In this way, the output from the vehicle soak time prediction model could be directly used as the input of MOBILE 6. Standard deviation would decrease significantly, which makes the distributions more stable.

5.1.2 Relationship between vehicle soak time and covariates

To test the variation of vehicle soak time distributions by other factors, we conducted statistical analysis on average vehicle soak time along time of day for several different groups.

The vehicle soak time distribution does not have the peak value at PM peak period on weekends, which is very reasonable. This weekday vehicle soak time distribution is very similar with weekend pattern. Also, in late hours of a day, the vehicle soak time duration will be longer than early hours, which means more people will stay in home after they leave their work place. On weekends, people have a longer vehicle soak time on the early morning because people usually have a long time for rest in the morning on weekends. A special phenomenon is that Sunday morning at 3 PM has high vehicle soak time than the other days in that time; it is possible that fewer people choose to leave at that time. It is may also be due to small sample size which may produce biased conclusion.

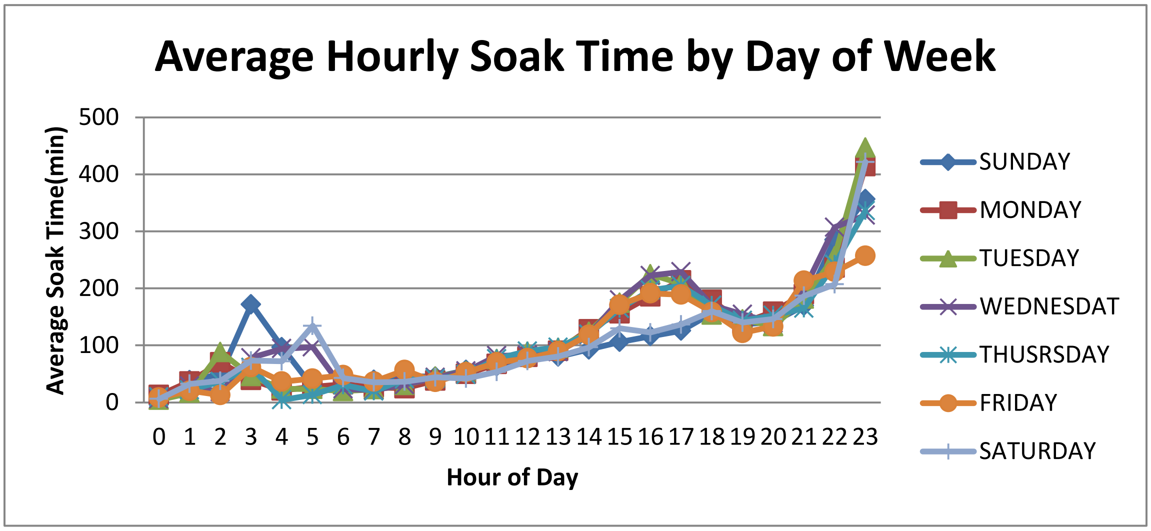


Figure 5-5: Distribution of Vehicle Soak Time by Day of Week

Generally, average vehicle soak time differ a lot by trip purpose. Home-based work trips would have much longer preceding vehicle soak time along the whole day. Noting that this is the vehicle soak time distribution for non-first start, we could find that people tend to stay at home for a relatively long period of time before going to work after other activities in the morning. And also home-based activities would have a longer vehicle soak time if they came back home later in the evening. In addition, home-based shopping has some zero values at the early morning because they will not choose to go shopping in the early morning. Home based shopping in the late night have the highest average vehicle soak time duration comparing with other activities; it is highly possible the sample includes some special days; people just stay at home for the day until the midnight and go out for special sales.

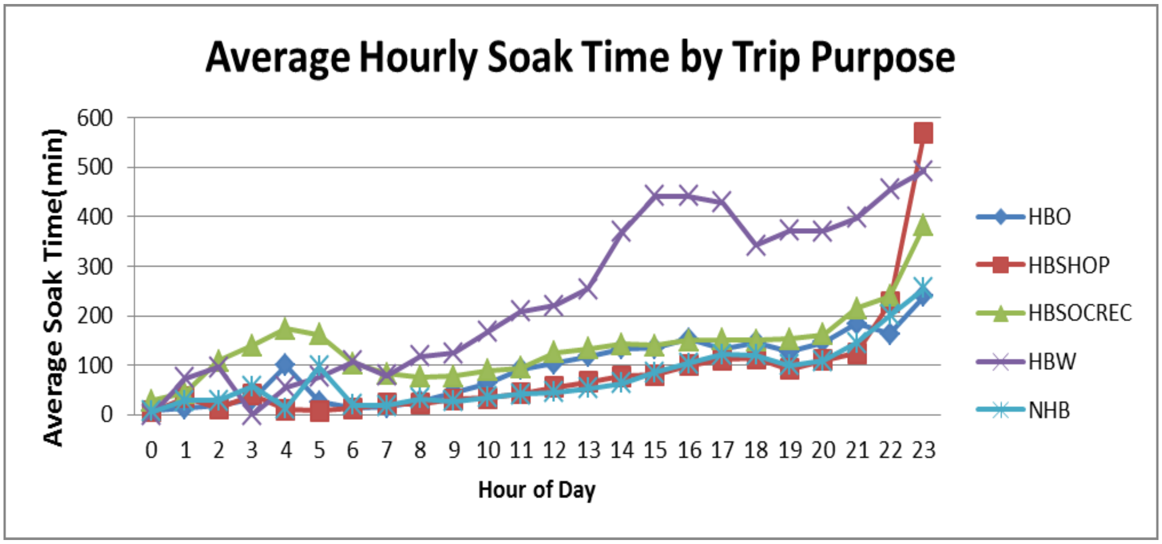


Figure 5-6: Distribution of Vehicle Soak Time by Trip Purpose

Figure 5-7 shows that vehicle soak time distributions for different vehicle types are slightly different. The major difference occurs at PM peak period. Pickup Truck has longer vehicle soak time and Van has shorter vehicle soak time than Car and SUV. The vehicle soak time duration for Pickup truck at 4am has a sudden peak; it is possible that people drive pickup trucks to some places where their work shifts starts 3am to 4 am. For the SUVs, it has a zero point at 4 am, which means nearly no people will be out at that time. Given the 2009 NHTS dataset, we have fewer data for other vehicle types, the graph for the average vehicle soak time by those kinds of vehicle types are biased. Thus, we did not show any analysis for the other vehicle types.

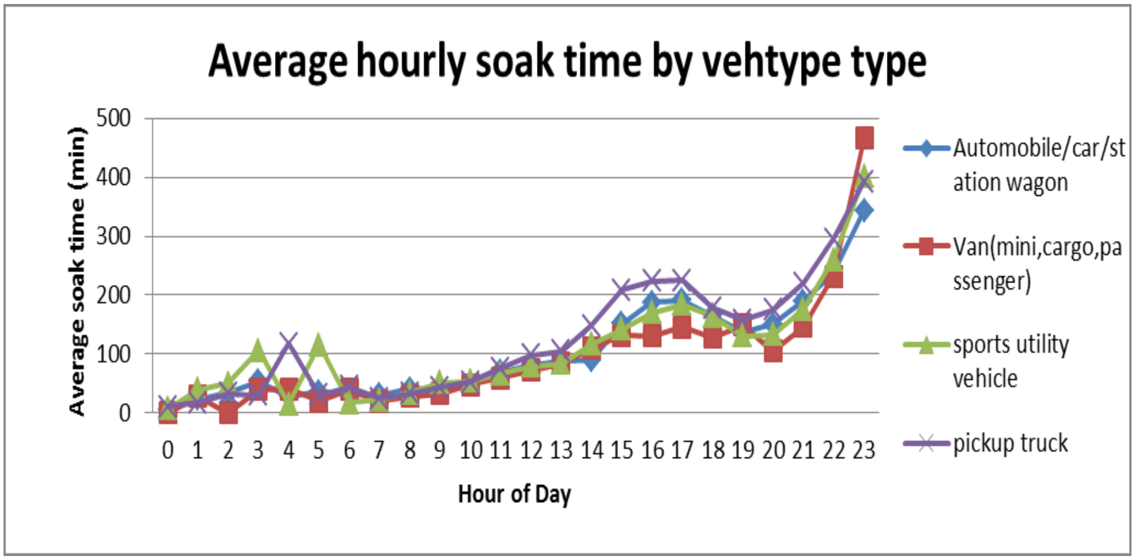


Figure 5-7: Distribution of Vehicle Soak Time by Vehicle Type

Besides the distribution by characteristics of trip and vehicle, we also conducted statistical analysis by attributes of household as well as regional characteristics.

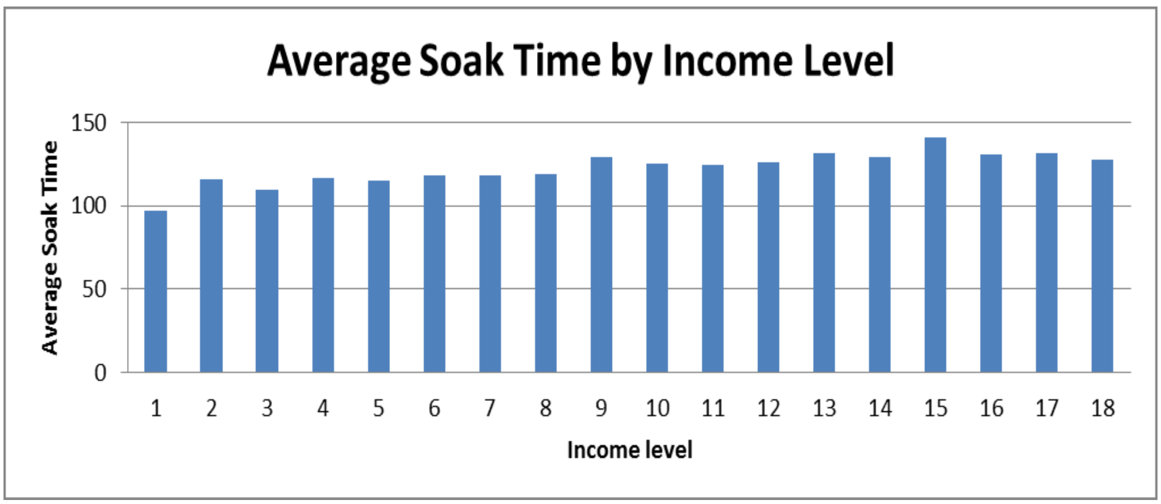
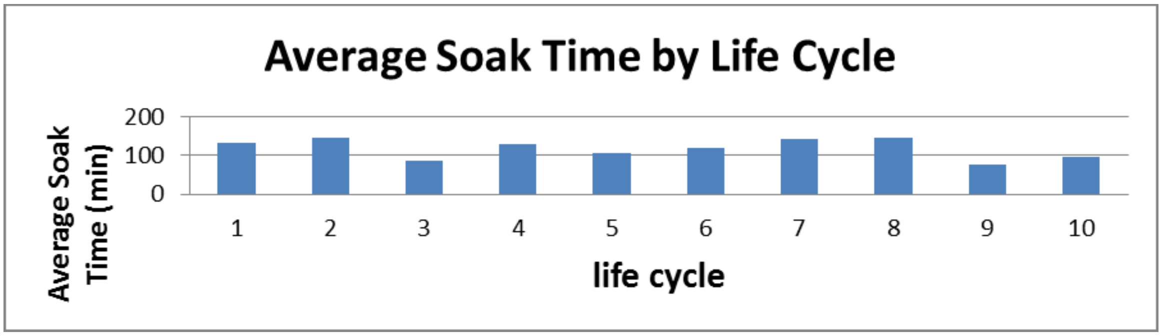
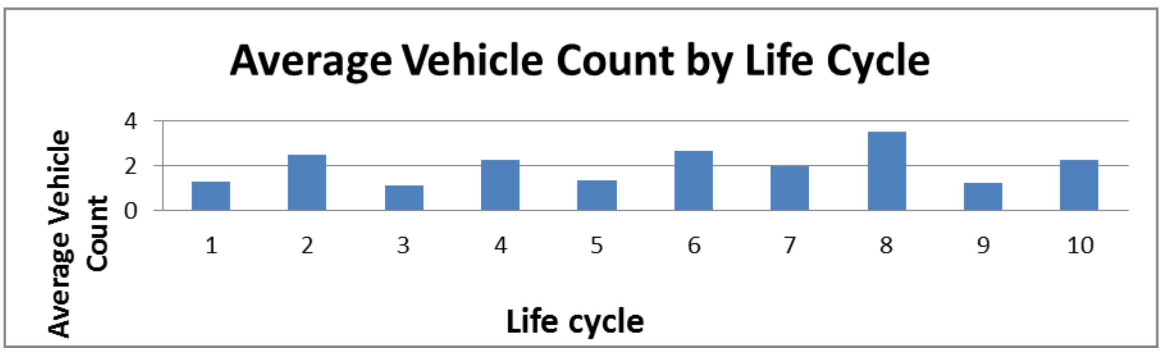


Figure 5-8: Distribution of Vehicle Soak Time by Household Income Level

Usually, the higher a household income is, the more vehicles a household haves. This phenomenon leads to fewer trips and longer average vehicle soak time for each household vehicle. However, the difference is very small, meaning the income level has minimum impact on vehicle soak time duration.

Average vehicle soak time duration by MSA size or Census Division have the same situation with the income level, which means the differences for different time of day are very minimum.





1=one adult, with no children	6=2+ adults, youngest children 6-15
2= 2+ adults, with no children	7= one adult, youngest children 16-21
3=one adult, youngest children 0-5	8=2+ adults, youngest children 16-21
4=2+ adults, youngest children 0-5	9= one adult, retired, on children
5= one adult, youngest children 6-15	10=2+ adults, retired, on children

Figure 5-9: Distribution of Vehicle Soak Time by Household Life Cycle

Lift cycle of a household has very significant influence on average vehicle soak time. Generally speaking, with the same number of children and age, households with two adults would make more trips every day, leading to lower vehicle soak times. However, just consider families with children, who are old enough to drive cars by their own, the number of vehicles for the household would increase dramatically. The average vehicle soak time for this kind of households is longer than those with young non-driving children. The average vehicle soak time for senior people is significantly lower. That's because senior people would not produce long vehicle soak period preceding work related trips. Instead, they would make more trips like shopping, going for leisure and so on with shorter vehicle soak time.

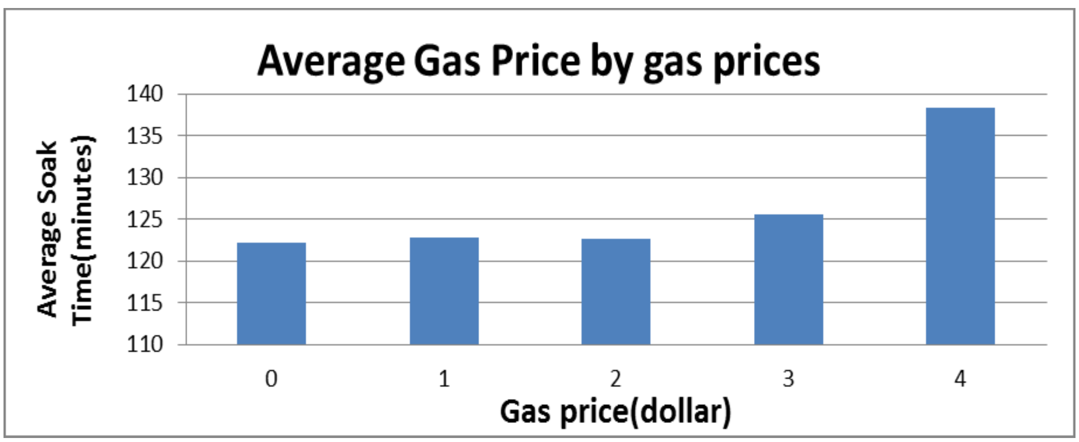


Figure 5-10: Distribution of Vehicle Soak Time by gas price

The most significant contribution in our research is the introduction of the gas price. High gas price leads to long vehicle soak time, meaning people combine their short trips into long ones. When the gas price is less than 3 dollars, the change for the vehicle soak time duration is not obvious.

5.2 Logistic Regression Model for First start

We introduce interaction variables among the time of day, the day of week, and the trip purposes to the logistic regression model.

Model Estimation

The logistic regression model results for first starts versus non-first starts are provided in Table 5-2. The base category used is non-first starts. A positive coefficient on the vehicle soak time duration indicates that this variable increases the probability of a first start, whereas a negative coefficient implies that the variable decreases the probability of a first start. The constant in the model does not have any behavioral interpretation.

The time of trip variables are introduced into the model to see the relationship between morning and weekend trips. The model results show that trips start occurring earlier in the day are more likely to be first starts. It also implies that it is less likely to see a weekend morning first start.

The activity-purpose variables are introduced into the model with the non-home based as the base purpose. The purpose dummy variables for most other home-based activities are positive, meaning if everything else is being equal, trips from home are most likely to be first starts. A comparison of the magnitudes of coefficients across all activity-purpose categories provides additional information about the likelihood of first starts among the group of home based trip starts. Specifically, home-based recreation trips and shopping trips are among those trips with highest possibility to be first starts. Work trips which are not so likely are due to the fact that a lot of work trips are observed to occur in the afternoon. We include two interaction variables of

HBW morning travel and HBR weekend travel, which indicates working trips in the morning are more likely to be first starts and recreation trips are fewer likely to be the first start (HBR weekend travel is not a significant contributor, though).

Table 5-2: Logistic Regression Model for First Starts

Variable	**coefficient**	P>z
Constant	-8.4744	
Vehicle Soak Time Duration	0.0114	0
Time of the Trip		
Morning Travel	3.1687	0
Weekend Travel	-0.0020	0.987
Weekend Morning	-0.1559	0.351
Activity purpose prior to trip start(Non-home purpose is base)		
Home-based Work(HBW)	0.1218	0.102
Home-based Shopping(HBS)	1.4732	0
Home-based Recreation(HBR)	1.8417	0
Home-based Other(HBO)	1.0173	0
HBW Morning Travel	1.4222	0
HBR weekend Travel	-0.5502	0.038
Number of Observations	405536	
log likelihood	-30244455	
Rho-Squared	0.8827	

5.3 Model for Vehicle Soak Time Duration for non-first Start

Regression analysis is often the chosen method to investigate impacts of a set of covariates on a dependent variable. However, regression requires that the dependent variable follows normal distribution. According to the statistics of non-first start vehicle soak duration, we could see clearly that the vehicle soak time duration preceding non-first starts is not normally distributed. Thus the generalized linear model (GLM) would be a more appropriate choice for our model establishment.

The model specification is formulated below.

$h(E(Y)) = h(\mu) = \alpha + \beta x + \gamma z$

Y denotes the response (or dependent variable)

x and z are two repressor variables (or predictors, covariates)

h is the link function)

The family option allows specifying the distribution and link function

Family=Gamma (link=Identity)

Table 5-3: Vehicle Soak Time Duration Model for non-first Starts

Variable	**coefficient**	**P**
Constant	-240.2247	0
Time of Vehicle Soaking(other time period is base)		
Early Morning	431.4262	0
Morning	258.1848	0
Time of non-first start trip(early morning and morning peak is base)		
Mid Day Travel	266.4154	0
PM peak	316.1093	0

Evening travel	369.8616	0
Night Travel	469.1753	0
Weekend Travel	5.3019	0
Weekend PM Peak Travel	2.1824	0.253
Activity purpose prior to the vehicle soaking period(other purpose is base)		
Work and School	7.5194	0
Work related	-1.2390	0.488
Recreation, medical/dental	3.7661	0.001
Return Home	7.7488	0
Shopping	0.6342	0.003
Serve Passenger	-6.1992	0
Lunch and dinner	-0.0939	0.931
Activity purpose of the non-first start trip(other purpose is base)		
Work and School	3.2299	0.017
Work related	6.1249	0.04
Recreation, medical/dental	1.5389	0.224
Return Home	1.7623	0.061
Shopping	1.8683	0.081
Serve Passenger	1.6761	0.366
Lunch and dinner	1.9167	0.139
Interaction variables(time of day and purpose of the prior trip)		
Early Morning return home	112.3025	0
Morning go to school/work	67.2051	0.002
Wkend medical and recreation	0.1327	0.949
wkend early morning return home	52.5577	0.328
Interaction variables(time of day and purpose of the non-first trip)		
PM peak return home	11.0235	0.01
Evening return home	-10.7523	0.065
Night go to work	4.1620	0.921
Weekend Travel	5.3019	0
Weekend PM Peak Travel	2.1824	0.253

Fuel variables		
gas price	2.0229	0.039
eiadmpg	0.1944	0.108
Gas price*mpg	-0.0682	0.122
Number of Observations	306446	
Rho-squared Stat	0.6	

The first step is to fit our sample to a parametric distribution. We tried to fit our data to four distributions, which are lognormal, gamma, exponential and inverse Gaussian distributions. Among the four distributions tested, Gamma distribution gave the smallest chi square value, indicating that Gamma distribution could best describe the sample among the four tested distributions.

The results for the vehicle soak time duration model for non-first starts are presented in Table 5-3. The dependent variable in the model is the vehicle soak time.

The time of the trip variables are introduced with the early morning and morning peak periods being the base. These two periods are combined into a single one because of very few non-first starts in these periods. The results for time of day variables indicate that the vehicle soak time preceding the non-first starts occurring later in the day is higher than for those occurring earlier in the day. And weekend trips tend to have shorter vehicle soak time for the reason that people make more trips during weekends. The time of vehicle soaking indicates that if people make a trip early in the morning, they tend to go for activities that would take a long time and thus lead to longer vehicle soak period.

The activity purposes are introduced with other purpose as the base activity. Interaction effects of activity purpose with time of the trip are also introduced. A great number of people have long vehicle soaking period after going to work or school. Thus we can observe the relatively higher coefficient for the work trips in the morning. We also observe that if people return home very early in the morning, they tend to have a sleep and get up late, and lead to a very long soak duration.

Another interesting finding is that if people go to work at night, they always stay home for rest and not go out during day time. Thus they have really long vehicle soak duration before going to work.

The most significant finding is that we introduce the gas prices and miles per gallon (MPG) data and the study of the relation between the vehicles soak time duration and gas price. The coefficient of the gas price is positive, meaning that with the increase of the gas prices, vehicle soak time duration will be much longer. It is highly possible that people combine their different short trips to long trips to decrease the travel time and save gas money, which definitely produce longer vehicle soak time. The interaction gas prices and MPG variables shows that even if gas price increases, people are more likely to choose the better MPG vehicle.

Same with the non-first start model, the first start model's results for time of the day variables indicate that the vehicle soak time preceding the non-first starts occurring later in the day is higher than for those occurring earlier in the day.

Trip purpose is also introduced in the first start model. It indicates that people have a long vehicle soak time after they finish the whole day activity. The return home purpose has the largest coefficient because people will have a long time for rest.

The negative coefficient of the gas price means when gas price increases, vehicle soak time duration decreases. It is highly possible that people change their departure time in the early morning to avoid congestion.

Chapter 6: Conclusions and future research

The multiple regression models with panel data analysis is developed to compare driver origin built environment factors and activity space level built environment factors impact on travel behavior, in particular person daily travel distance. The model allows analysts and decision makers to estimate the VMT reduction effects of various proposed built environment changes, such as higher residential density, employment density, land use mixed development, smaller block size and so on.

Our findings show that compact, mixed land use and smaller block sizes can be effective in reducing VMT per person at daily level. These measures can help reduce traffic congestion, reduce energy consumption, and enhance environmental quality.

Given the Maryland GPS data's limitations on its sample size, it is highly challenging to accurately and quantitatively attribute the socio-demographic impacts on VMT. However, the type of such GPS data which offer precise location information for each person at each minute enables us accurately capture of detailed person travel information. When matched with the land use data, the TAZ level land use can be obtained for the whole travel trip.

Based on the distribution test analysis on vehicle soak time, it is observed that a number of attributes have significant impact on vehicle soak time distribution. Such attributes include trip purpose, time-of-day, vehicle type, and life cycle.

The logistic regression model of start patterns interprets the behavioral pattern for first and non-first starts. It indicates there is significant difference between first start trips and non-first start trips. The treating these two types of trips independently in the vehicle soak duration model is needed.

And lastly, the vehicle soak-time duration has been modeled by using a generalized gamma model, with time-of-day, purpose, and some interaction variables specified. It is worth noting that the rho-squared statistic of the duration model has been significantly undermined by the relatively high data variance and constraint of variable distribution, which also provides an insight where our future research endeavor lies.

We introduce gas prices into all trips model. The model explains that people are now paying more attention on the relation between emission and fuel consumption. With the high gas price, people are more likely to adjust their travel arrangement and choose some high mpg vehicles as the travel tool. Moreover, the choice for the departure time and season change also bring some possible impact on vehicle soak time duration.

References:

An, F., M. Barth, G. Scora, and M. Ross. "Modal-Based Intermediate Soak-Time Emissions Modeling." In Transportation Research Record: Journal of the Transportation Research Board, No. 1664, TRB, National Research Council, Washington, D.C., 1999, pp. 58–67.

Assessment and Standards Division Office of Transportation and Air Quality U.S. Environmental Protection Agency. "User's Guide to MOBILE6.1 and MOBILE6.2". August 2003.

Brownstone, D. (2008). Key Relationships Between the Built Environment and VMT.

Boarnet, M. G., Houston, D., Ferguson, G., & Spears, S. (2010). Land Use and Vehicle Miles of Travel in the Climate Change Debate: More than Your Average Bear.

Brüderl, J. (2005). Panel Data Analysis.

Boarnet, M. G., & Crane, R. (2001). The influence of land use on travel behavior: specification and estimation strategies. Transportation Research Part A, 35(9), 823-845.

Boarnet, M. G., & Sarmiento, S. (1998). Can Land-use Policy Really Affect Travel Behavior? A Study of the Link between Non-work Travel and Land-use Characteristics. Urban Studies, 35(7), 1155-1169.

Bottai, M., Salvati, N., & Orsini, N. (2006). Multilevel models for analyzing people's daily movement behavior. Journal of Geographical Systems, 8(1), 97-108.

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2006). Neighborhood design and vehicle type choice: Evidence from Nor thern California. Transportation Research Part D, 11(2), 133-145.

Cervero, R. (1991). Land Uses and Travel at Suburban Activity Centers. Transportation Quarterly, 45(4), 479-91.

Cervero, R. (1996). Mixed land-uses and commuting: Evidence from the American Housing Survey. Transportation Research Part A, 30(5), 361-377.

Cervero, R. (1998). Land-use mixing and suburban mobility. Transportation Quarterly, 42(3), 429-46.

Cervero, R., & Kockelman, K. (1997). Travel Demand and the 3Ds: Density, Diversity, and Design. Transportation Research Part D, 2(3), 199-219.

Chen, C., Gong, H., & Paaswell, R. (2008). Role of the Built Environment on Mode Choice Decisions: Additional Evidence on the Impact of Density. Transportation, 35, 285-299.

Cities, T. U. S., Zhang, L., & Author, C. (n.d.). How Built Environment Affects Travel Behavior : A Comparative Analysis of the Connections between Land Use and Vehicle Miles, 1–19.

Dargay, J. M., & Hanly, M. (2003). The Impact of Land Use Patterns on Travel Behaviour, *44*(October), 0–15.

Duch, R. (2008). Longitudinal / Panel Data Analysis, 1–24.

Ewing, R., and R. Cervero. Travel and the Built Environment: A Synthesis. (2001).

EPA. (2006). Greenhouse Gas Emissions from the U.S. Transportation Sector, 1990-2003.Washington, DC: Environmental Protection Agency.

Ewing, R. and Cervero, R., 2010. Travel and the Built Environment; A Meta-Analysis. Journal of the American Planning Association, vol. 76, No. 3, 265-295

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis. Transportation Research Record, 1780, 87-113.

Effects, R. (n.d.). Panel Data Analysis Fixed & Random Effects, 1–40.

Ewing, R., Bartholomew, K., Winkelman, S., Walters, J., & Chen, D. (2008). Growing Cooler. Washington, DC: ULI-the Urban Land Institute.

Frank, L. D. (2000). Land Use and Transportation Interaction: Implications on Public Health and Quality of Life. Journal of Planning Education and Research, 20(6), 6-22.

Frank, L. D., Saelens, B. E., Powell, K. E., & Chapman, J. (2007). Stepping towards causation: Do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity? Social Science & Medicine, 65, 1898-1914.

Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, T. K. (2008). Urban form, travel time, and cost relationships with tour complexity and mode choice. Transportation 35(1), 37-54.18

Frumkin, H. (2002). Urban sprawl and public health. Public Health Reports, 117, 201-217.

Frank, L. (2005). CORRELATION ANALYSIS.

Gao, H. O. and L. S. Johnson (2009). "Methods of analysis for vehicle soak time data." Transportation Research Part A 43(2009): 744-754.

Glover, E. L. and P. Carey. “Determination of Start Emissions as a Function of Mileage and Soak-Time for 1981–1993 Model Year Light-Duty Vehicles." Report No. M6.STE.003. Environmental Protection Agency, Office of Mobile Sources, Assessment and Modeling Division, 1999.

Golberg, M.A., Cho, H.A., 2004. "Introduction to Regression Analysis." WIT Press, Southampton, United Kingdom.

Gelman, A., & Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. New York: Cambridge University Press.

Gelman, A., & Pardoe, I. (2006). Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models. TECHNOMETRICS, 48(2), 241-251.

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment*, *10*(6), 427–444.

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern California. Transportation Research D, 10, 427-444.

Hanson, S. and Genevieve G. (2004). The Geography of Urban Transportation, 3rd Edition. The Guilford Press.

Irvine, U. C., & Sarmiento, S. (1996). Can Land Use Policy Really Affect Travel Behavior? A Study of the Link Between Non-Work Travel and Land Use Characteristics.

Kitamura, R., Mokhtarian, P. L., & Daidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. Transportation, 24(2), 125-158.

Koupal, J., M. Cumberworth, et al. (2002). Draft Design and Implementation Plan for EPA's Multi-Scale Motor Vehicle and Equipment Emission System (MOVES).

Krizek, K.J. (2003a). Residential relocation and changes in urban travel: Does Neighborhood-Scale Urban Form Matter. Journal of the American Planning Association, 69(3), 265-281.

Krizek, K.J. (2003b). Neighborhood services, trip purpose, and tour-based travel. Transportation, 30(4), 387–410.

Levinson, D. and K.J. Krizek (2008) Planning for Place and Plexus: Metropolitan Land Use and Transport.

Land, H., Factors, U., Travel, A., & Litman, T. (2011). Land Use Impacts on Transport.

McMullen, B.S., Zhang, L., and Nakahara, K. (2010). Distributional Impacts of Changing from a Gasoline Tax to a Vehicle-Mile Tax. Transport Policy 17(6): 359-366.

Mokhtarian, P. L., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. Transportation Research B, 42 (3), 204-228.

National Academy of Science (2009). Driving and the Built Environment: The Effects of Compact Development on Motorized Travel, Energy Use, and CO2 Emissions. Available online at: http://www.nap.edu/catalog/12747.html

Neuman, M. (2005). The Compact City Fallacy. Journal of Planning Education and Research,25, 11-26.

Nair, H. S., C. Bhat, et al. (2002). "Modeling soak-time distribution of trips for mobile source emissions forecasting: techniques and applications." Transportation Research Record (1750): 24-31.

National Research Council (NRC), 2004. "Air Quality Management in the United States. " National Academy Press, Washington, DC.

Panel Data Analysis October 2011. (2011).

Paper, D., Vol, S., & Development, I. (2006). Land Use Effects on Travel Behavior in Jabotabek (Indonesia) Metropolitan Area Metin SENBIL, Junyi ZHANG, Akimasa FUJIWARA, *2006*.

Pushkarev, B. S., and J. M. Zupan (1977). Public Transportation and Land Use Policy. Bloomington: Indiana University Press. Transit Cooperative Research Program (TCRP).

Rubin T, Lowe M, Aytekin B, Gereffi G (2010) Does Bus Transit Reduce Greenhouse Gas Emissions? Retrieved from http://reason.org/news/show/does-bus-transit-reduce-greenhouse

Silva, J. D. A., Golob, T. F., & Goulias, K. G. (2006). Effects of Land Use Characteristics on Residence and Employment Location and Travel Behavior of Urban Adult Workers. *Transportation Research Record: Journal of the Transportation Research Board*, *1977*(-1), 121–131.

Southworth, F. (2008). On the potential impacts of land use change policies on automobile vehicle miles of travel, *29*(2001), 1271–1283.

Schwanen, T., & Mokhtarian, P. (2005). What affects commute mode choice: neighborhood physical structure or preferences toward neighborhoods? Journal of Transport Geography,13, 83-99.

US Environmental Agency (EPA), 1994. "User's guide to Mobile5 (mobile source emission factor model)," EPA-AA-AQAB-94-01.

US Environmental Agency (EPA), 1994. "User's guide to Mobile 6.1 and Mobile 6.2(mobile source emission factor model)," EPA420-R-03-010.

Venigalla, M. M. and D. H. Pickrell (2002). "Soak distribution inputs to mobile source emissions modeling: measurement and transferability." Transportation Research Record (1815): 63-70.

Venigalla, M., T. Miller, and A. Chatterjee. "Start Modes of Trips for Mobile Source Emissions Modeling." In Transportation Research Record 1472, TRB, National Research Council, Washington, D.C., 1995, pp. 26–34.

Washington D.C., EPA420-P-02-006, US Environmental Protection Agency, Office of Transportation and Air Quality.

Zegras, C. (2004). Influence of Land Use on Travel Behavior in Santiago, Chile. *Transportation Research Record*, *1898*(1), 175–182.

Zegras, C. (2010). The Built Environment and Motor Vehicle Ownership and Use: Evidence from Santiago de Chile. Urban Studies, 47, 1793-1817.19

Zhang, L. (2010). Do freeway management strategies exacerbate urban sprawl? Journal of the Transportation Research Board 2174: 99-109.

Zhang, L., McMullen, B.S., Nakahara, K., Valluri, D. (2009). The short- and long-run impact of a vehicle mileage fee on income and spatial equity. Journal of the Transportation Research Board 2115: 110-118