Estimating and Modeling Soak Time Distributions with the 2009 National Household Travel Survey Data

Lei Zhang[1,*], Xiaojie Cong[2], Xiang He[2], and Chenfeng Xiong[2]

1. Assistant Professor

2. Graduate Research Assistant

Department of Civil and Environmental Engineering

University of Maryland

1173 Glenn Martin Hall, College Park, MD 20742

Phone: 301-405-2881

Email: lei@umd.edu

* Corresponding Author

Draft: July 5th, 2011

Word Count: 4925 + 13 Tables and Figures = 8275

Abstract

Vehicle soak time is defined as the duration of time a vehicle's engine is at rest prior to being started. The distribution of soak time is a key input for mobile-source emission models, such as the EPA MOVES. This paper estimates various soak time distributions, and develops statistical models of those distributions.

The data source is the National Household Travel Surveys (NHTS) in 2009, which contains information about person and vehicle trips in a 24-hour period for all sampled households. When the weights are introduced, the total vehicle trips for the national level are 467,505,568. We first develop a comprehensive methodology for extracting vehicle soak time distribution information from NHTS data files including the day trip file, person file, vehicle file and household file. The obtained soak time information is then employed in the development of a series of statistical models that can directly provide inputs to mobile-source emission models.

Vehicle emission rates are heavily influenced by soak time distributions due to their impact on vehicle start emissions and evaporative emissions. Since the distribution and duration of soak periods preceding the first vehicle start of day is quite different from those of soak periods preceding non-first starts, we analyze these two types of soak period with separate models. Results show that time of day, day of week, trip purpose, vehicle type, gas price, metropolitan statistical area size and several interaction variables have significant impacts on soak time durations.

A model analyzing the start mode fraction is built with logistic regression methods. The model rho-squared is 0.88 based on more than 0.41 million observations. Again, time of day, trip purpose, day of week and their interactions are found to be the main factors explaining the differences between soak periods prior to first start and those of non-first starts. Following the start mode fraction model, a statistical model on non-first start soak time durations is also established. After fitting the data with several parametric distributions, the generalized Gamma model is chosen for its superior goodness of fit. This model enables emission modelers and analysts to predict soak time distributions based on several demographic, socioeconomic and travel behavior characteristics. The impact of fuel price on soak time is also considered in the model.

1. Introduction

As defined in those emission models, soak time means the duration of time in which the vehicle's engine is not operating and that precedes a successful vehicle start, which determines the duration of the evaporative emissions. Therefore, exploring the distribution of soak time and how it can be modeled and predicted is of particular interest among decision makers and researchers. Gas price as a new variable is also introduced here, which help to furthermore interpret the relationship between some emission variables and soak time distributions.

The data source is the National Household Travel Surveys (NHTS) in 2009, which contains information about person and vehicle trips in a 24-hour period for all sampled households. When the weights are introduced, the total vehicle trips for the national level are 467,505,568, which offer the more accurate data sources for the research.

What is more, soak time has also been correlated with the vehicle start information, such as whether a vehicle start is cold (longer soak time duration) or hot (shorter soak time duration), and whether it is a first start or a non-first start. These variations result in large difference in vehicle start emission factors, which also needs careful measurement. Thus, a desirable analysis tool for soak time and its distribution can better clarify the relationship between the length and corresponding vehicle start characteristics of a soak period, and hence provide sound input to the emissions models.

The goal of this paper is to conduct a pilot study on soak time related topics using the NHTS 2009 data. Firstly, the authors conduct a series of distribution studies of the soak time data. Reflected by the new data, some new interpretations are crucial, especially when the new national travel survey has involved more information about the household. The study of distributions is worthwhile and can therefore provide an insight about how to model the soak time distribution properly in the future research. Secondly, the paper conducts an empirical study about one of the key issues in vehicle soaking and starting, the start mode fractions. The model is built on binary logit structure, with some improved modeling specifications. Last but not the least, the application capability of the soak time studies coupled with the future research scopes have been extensively interpreted, which as well leads to our most short-run research goals.

The remainder of the paper is organized as follows. In Section 2, available literatures have been scanned and hence the paper comes up with the basic research ideas. In the following section, the data availability and model specification are introduced. The results are presented in the Section 4. Section 5 concludes the paper and discusses the model application and future research work.

2. Literature Review

While being relatively new, this field has been explored by a couple of researchers. These early 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, using simply the time-of-day and trip purpose variables. Some later papers advanced this model specification considerably by incorporating some zonal attributes. Thirdly, these studies all have huge connection to the EPA 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 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 claimed by this paper closely correlated to 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 soak distribution inputs by time period. The grouping scheme will enhance the utility of survey data in deriving the soak distributions and reduce the effort in providing soak distribution inputs to MOBILE6.Nair et al. (2002) also studied the data from the 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., but focused on modeling the disaggregate soak-time distributions, which are important input of the MOBILE6.

Gao and Johnson (2009)'s paper has 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 paper is finalized by a general guideline for the analysis of soak time data. Finally, a subset of the statistical methods discussed is used to analyze the US Environmental Protection Agency's 3-city data.

Glover and Carey's paper uses three Federal Test Procedure (FTP) data sources: 1) the test results from the EPA laboratory in Ann Arbor, Michigan, 2) the data received from AAMA (American Automobile Manufacturers Association) based on testing conducted in Michigan and Arizona, and 3) the API (American Petroleum Institute) data collected in Arizona. The dataset combines vehicle type, model year and technology together. In their research, start emissions are a function of soak time so that MOBILE6 will be able to account for the entire distribution of soak times observed in the fleet. Soak time ranges from zero minutes up to a 12 hour soak period (720 minutes). Soak periods exceeding 12 hours will be assumed to be the same as for a 12 hour soak in this model.

Venigalla and miller classify the trip start as a cold start and a hot start, the emission modeling process uses 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 exercises.

MOBILE 6 is a computer program that estimates some emission factors for gasoline-fueled and diesel highway motor vehicles, and for some specialized vehicles. The input parameters for the MOBILE6 include calendar year, month, and vehicles characteristics; especially include engine starts soak time distribution by hour, which is our focusing point in the research. The output will include emissions from hot soak conditions, running 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 the soak time duration related variables. The MOVES also includes spreadsheet tools that help with the transition from MOBILE6.2 to MOVES.

In this paper, as in the paper by venigalla and Pickrell, national level travel survey data are used. However, there are important differences between this approach and that of Venigalla and Pickrell. First, the focus of this study is the disaggregate soak time distributions, whereas that of Venigalla and Pickrell (2002) was the more aggregate start mode fractions. Second, this study examines the start pattern as a function of a series of interaction variables in addition to time of day and purpose, which is proved to be more accurate. Therefore, in the third place, soak-time distributions for first starts and nonfirst starts of the day are explicitly distinguished because the distributions for these two types of starts are likely to be very different. Just estimating and modeling the soak time distributions, our paper have very similar research procedure with Gao and Johnson. Firstly, the paper focuses on the difference of first and non-first start soak time distribution; then give the soak time distribution considering the time of day; in the following part, study the relationship between soak time duration and some corresponding variables, finally modeling the soak time distributions, using the regression model to see whether a trip is a first start trip or non-first start trip, the generalized liner model to analyze the dependent soak time duration variable by first start and non-first start separately.

3. NHTS Data Processing

The National Household Travel Survey with the weighted data is a household-level survey conducted all over 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 gas prices file at month level for different states from 1990 to 2011.

For our research purpose, the travel day file would be the primary source to compute soak time, as the information for each trip is recorded row by row. In addition, by keeping a unique household ID, vehicle ID and person ID for each soak period, we would be able to join the soak time file we got with the household file, vehicle file and person file. This would allow us to conduct analysis on the impact of household, vehicle and person characteristics on soak time duration and distributions.

However, there are some issues relating to the travel day file, which would cause some problems when computing 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 several steps.

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 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 a part of the trips 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 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 soak period preceding that trip or following that trip. Thus we delete all the trips made by that vehicle.

Step IV: Deal with the overlapping of trip time and compute 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 2.

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

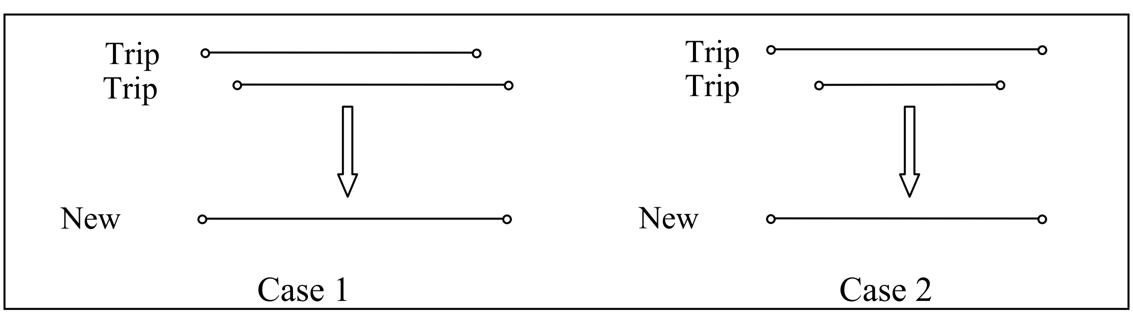


Figure 2. Overlapping of Trips

As the NHTS only collects data in a 24-hour period, determining the 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 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.

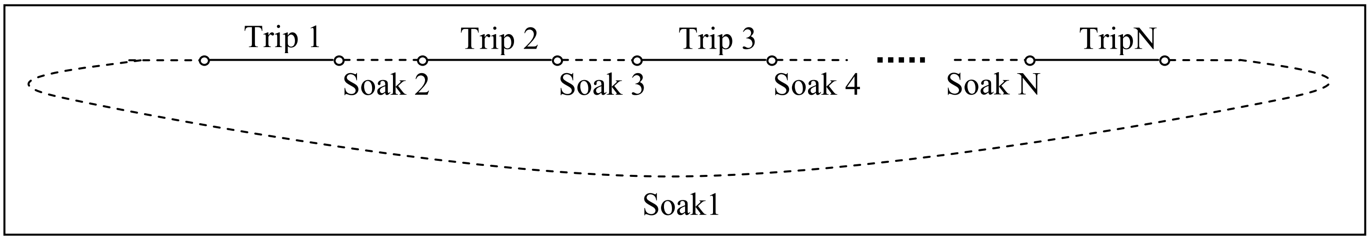


Figure3. Soak Time along a Whole Day

Step V: Delete the vehicles which have only one 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 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 soak time more realistic, we'd rather delete all the vehicles with only one soak time record.

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

For the purpose of exploring soak time distribution by different characteristics, we need to join the derived 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 soak time records.

4. Soak Time Distributions

4.1 Soak time distributions for first start and non-first start

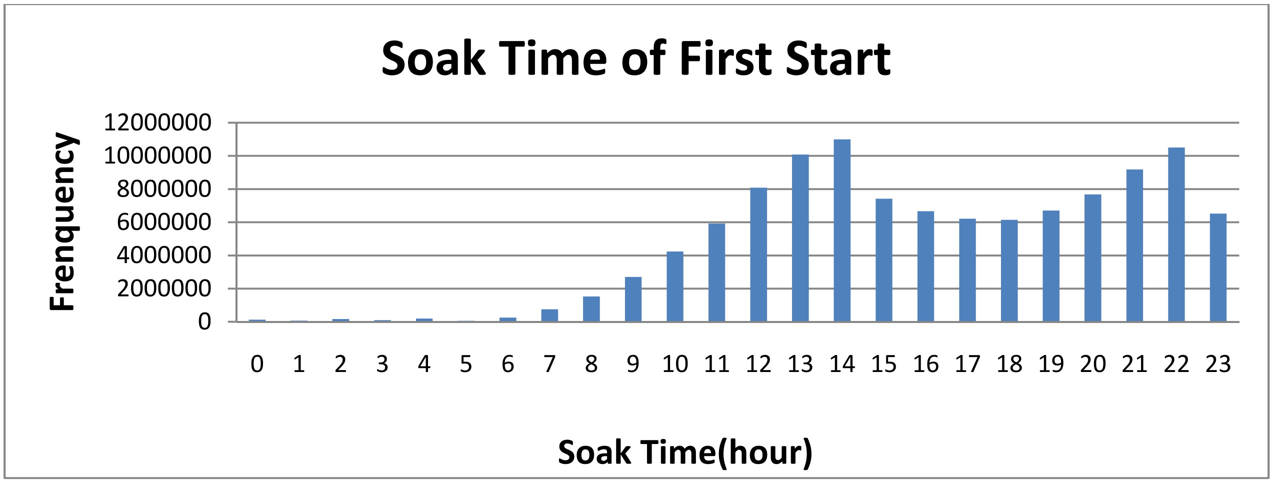
From our common sense, soak period preceding first starts should be relatively long compared to soak periods preceding non-first starts. Meanwhile, as the distribution of first starts and non-first starts by time of day would be significantly different, the distribution of soak times for first and non-first starts would also differ a lot along the time of day.

Figure 1 shows the distributions of duration of soak time for first starts and non-first starts. Based on the current EPA model MOBILE6.2, a cold start is defined as a successful vehicle start following a 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 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 2. In addition, it would influence the average soak time and the standard deviation of soak time as well.

The average soak time for the non-first starts during PM peak period is relatively high, which represents the activities of leaving work place for lunch from several hours morning work. After they finish their lunch, it will have a long time to stay in the office or studying places in the afternoon, therefore, soak time durations are much longer in the 5 pm until they leave after finishing the day's work. The long soak time at late night represents the activities of going back home from work or pleasure. People will not go out more often in the late night.



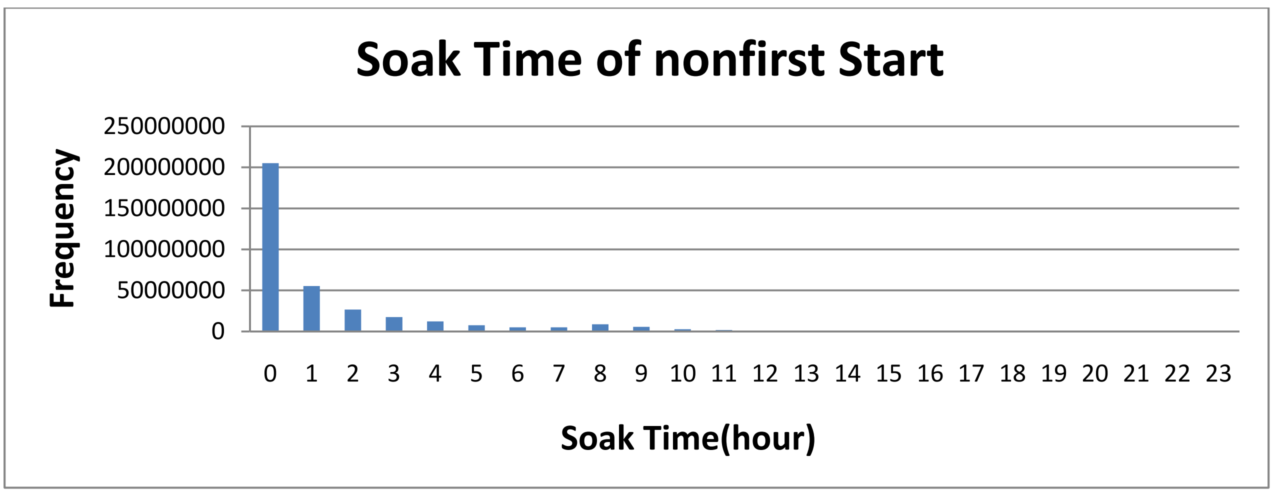


Figure 1. Distributions of Soak Time Duration

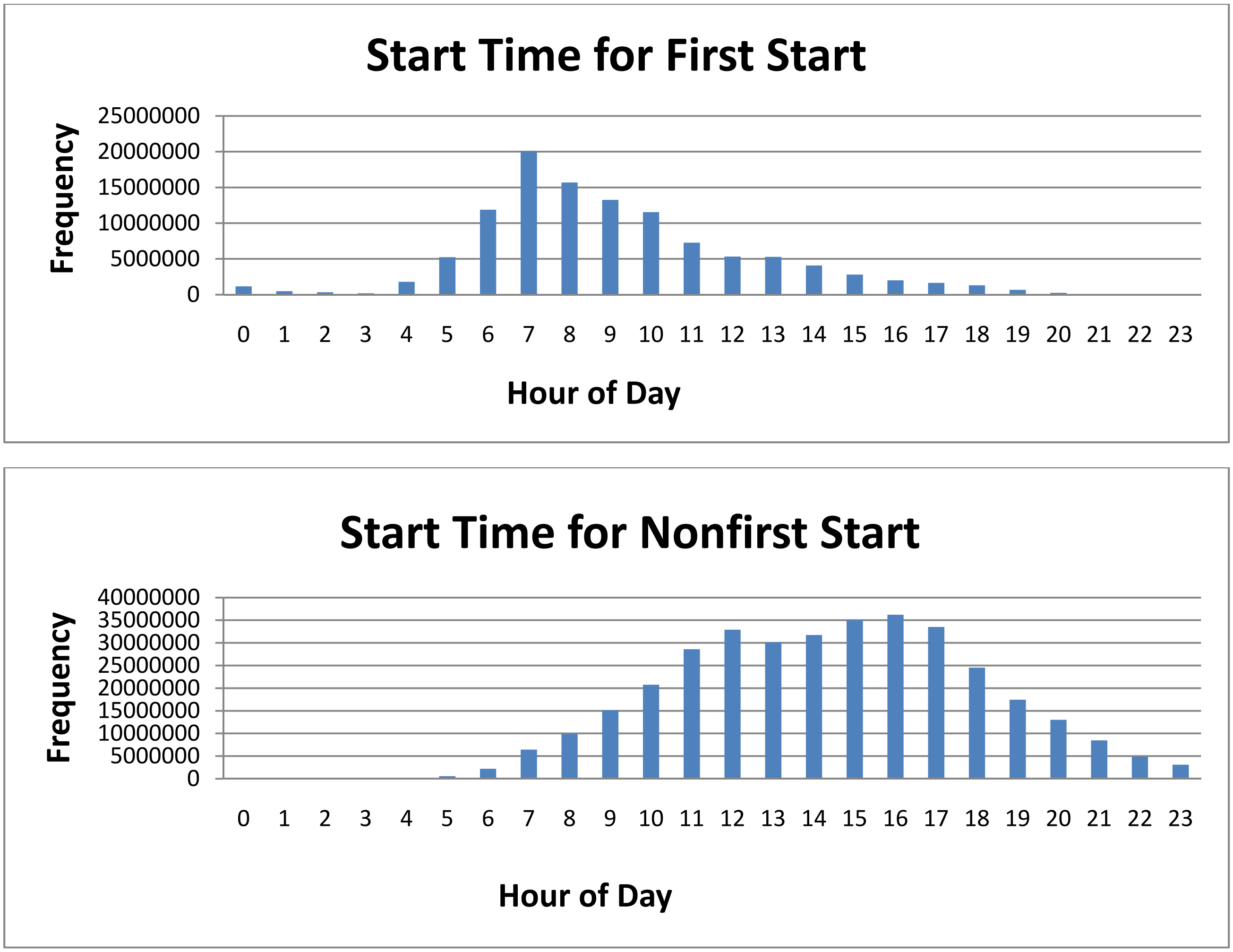


Figure 2. Distribution of Vehicle Start Times

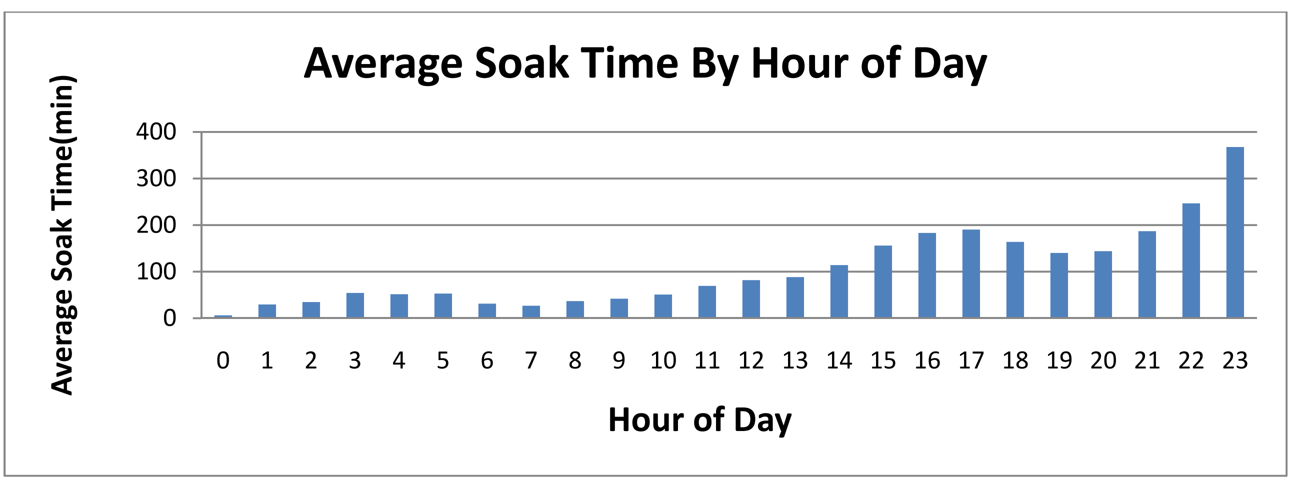


Figure 3. Distribution of Average Soak Time by Time of Day

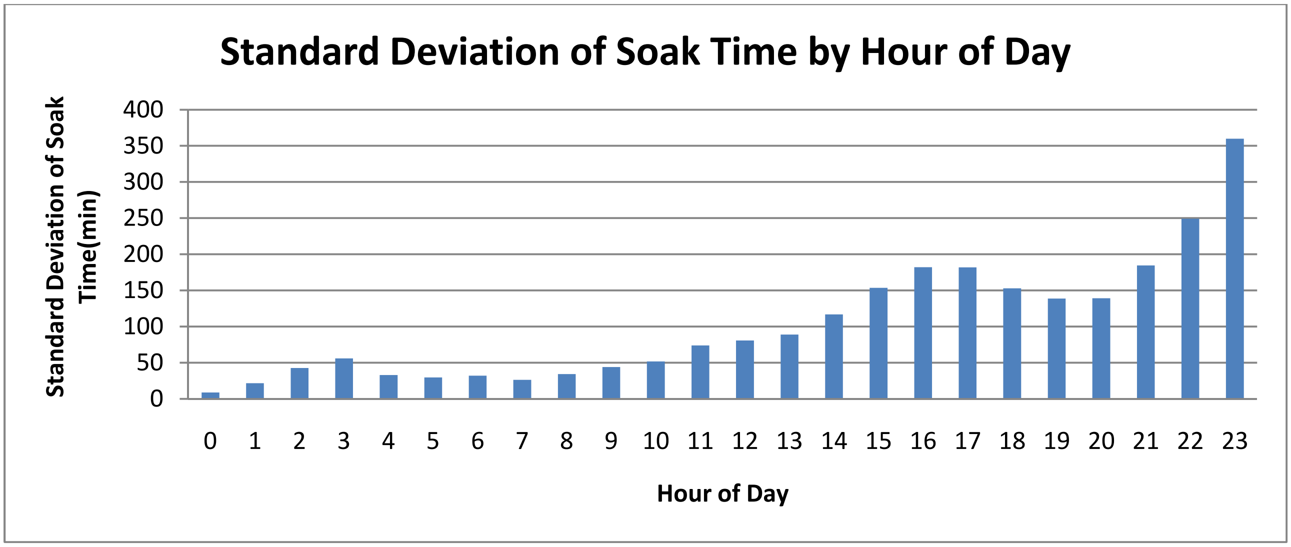


Figure 4. Distribution of Standard Deviation of Soak Time by Time of Day

A potential application of the analysis of soak time distributions is to derive the inputs for some existing mobile source emissions models. By comparing Figure 3 and Figure 4, we could find that the ratio of average soak time to the standard deviation of soak time along hour of day is very close to 1 and in a lot period the ratio is lower than 1, which means that the soak time distributions are not stable along time of day. To improve the quality of prediction, two sets of method could be implemented simultaneously.

1. Incorporate other factors besides time of day when predicting soak time distribution.

Although average 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 soak time. These factors may contribute a lot in the deviations.

2. Divide the soak time into time bins.

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

Table 1. MOBILE 6 Soak Time Input Format

Soak Interval (N)	Range of 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 prediction of soak time could be directly used as the input of MOBILE 6, and the standard deviation would decrease a lot, which makes the distributions more stable.

4.2 Relationship between soak time and covariates

To test the variation of soak time distributions by other factors besides time of day, we conducted some statistics on average soak time along time of day for several different groups.

The soak time distribution does not have the peak value at PM peak period on weekends, which is very reasonable. And the soak time distribution is very similar for weekdays and for weekends. Also, in the late day, the soak time duration will be longer than the early time, which means more people will stay in home after they leave their work place. On weekends, people will have a longer soak time on the early morning because people usually have a long time for rest in the morning on weekends. The special point is that Sunday morning at 3 PM has an obvious high soak time than the other days in that time; it is possible that fewer people will chose to leave at that time, so the sample size is very small that produces the biased situation.

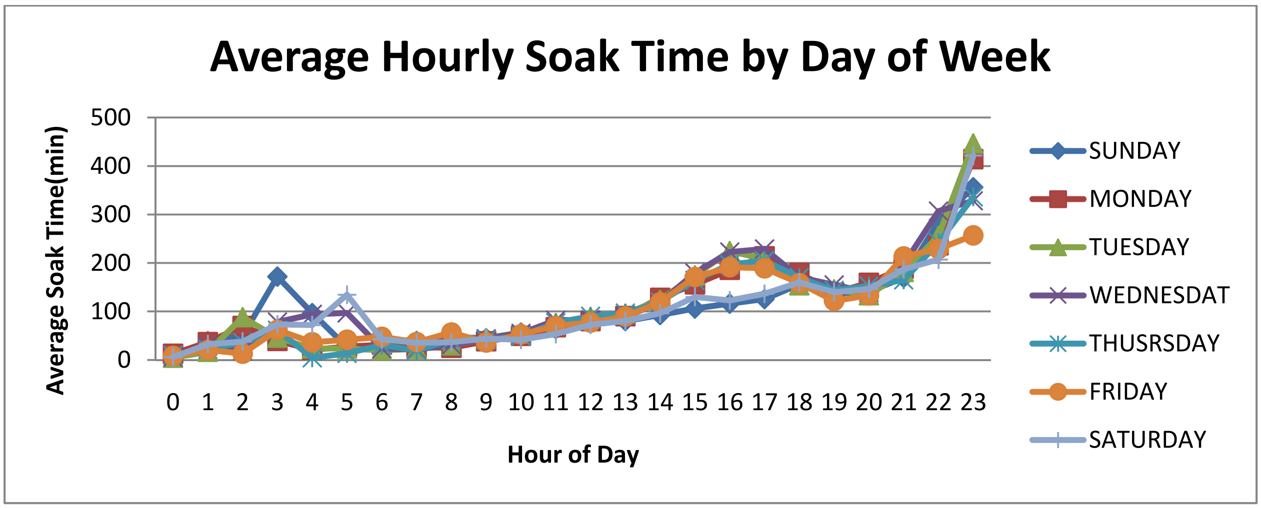


Figure 5. Distribution of Soak Time by Day of Week

Generally, average soak time differ a lot by trip purpose. Home-based work trips would have much longer preceding soak time along the whole day. Noting that this is the 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 soak time if they went back home later in the evening. In addition, home-based shopping will have some zero values at the early morning because they will not choose to go shopping in the early morning. Moreover, home based shopping in the late night will have the highest average 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 some sales commodities.

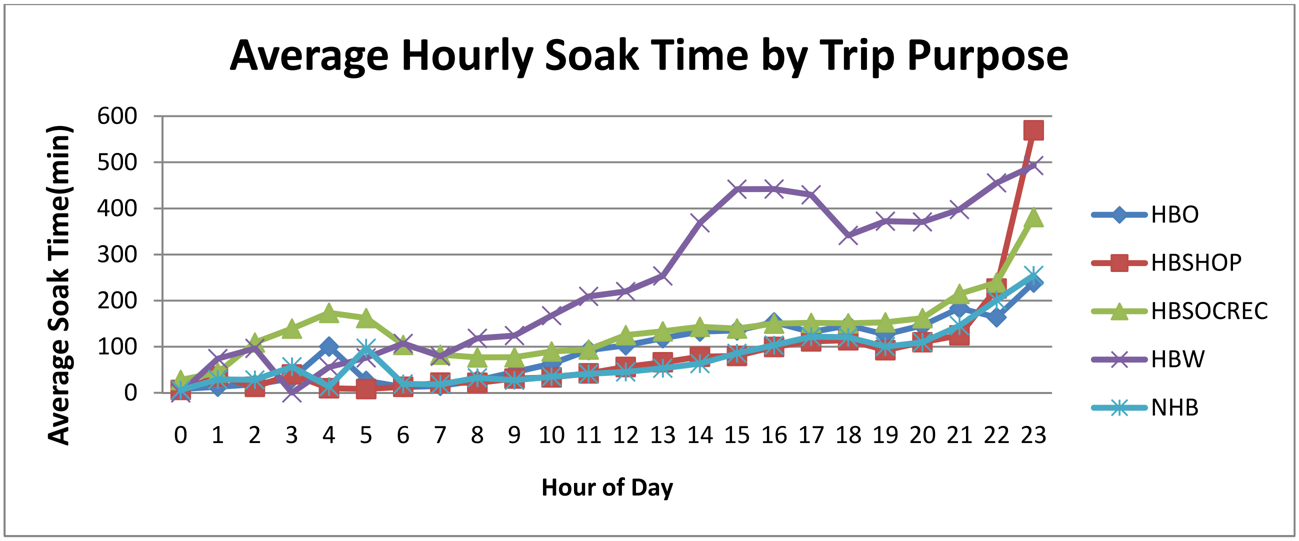


Figure 6. Distribution of Soak Time by Trip Purpose

Figure 7 shows that soak time distributions for different vehicle types are slightly different, and the major difference occurs at PM peak period. Pickup Truck has longer soak time and Van has shorter soak time than Car and SUV, which are usually used as commuting vehicles. The soak time duration for Pickup truck at 4am has a sudden peak; it is possible that people will drive the pickup truck to some place and stop, then they will start to work at 3am to 4 am. For the SUV, it has a zero point at 4 am, it means nearly no people will be out at that time, and more likely, and then sample size is very small, so the result is a little biased at the early morning time. However, in the 2009 NHTS dataset, we have little data for the other vehicle types, the graph for the average soak time by those kinds of vehicle types are biased. Thus, we did not show any analysis for the other vehicle types.

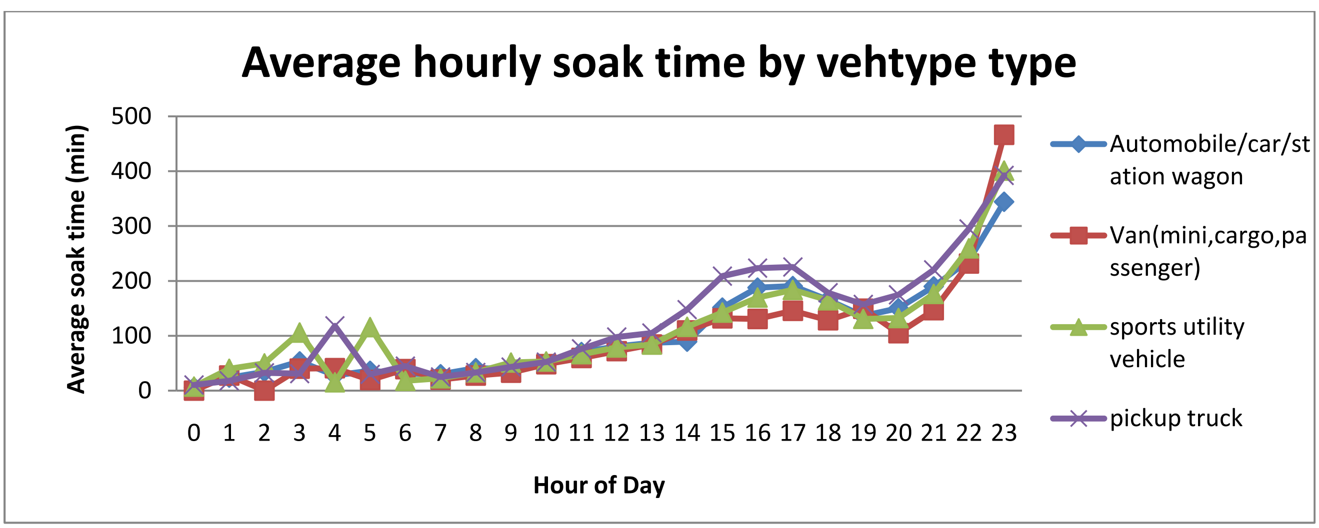


Figure 7. Distribution of Soak Time by Vehicle Type

Besides the distribution by characteristics of trip and vehicle, we also conducted statistics by attributes of household as well as regional characteristics. Usually, the higher the household income is, the more vehicles the household would have, which leads to the fewer trips and longer average soak time for each household vehicle. But the difference is very small; it means the income level will impact the soak time duration little.Average 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 small.

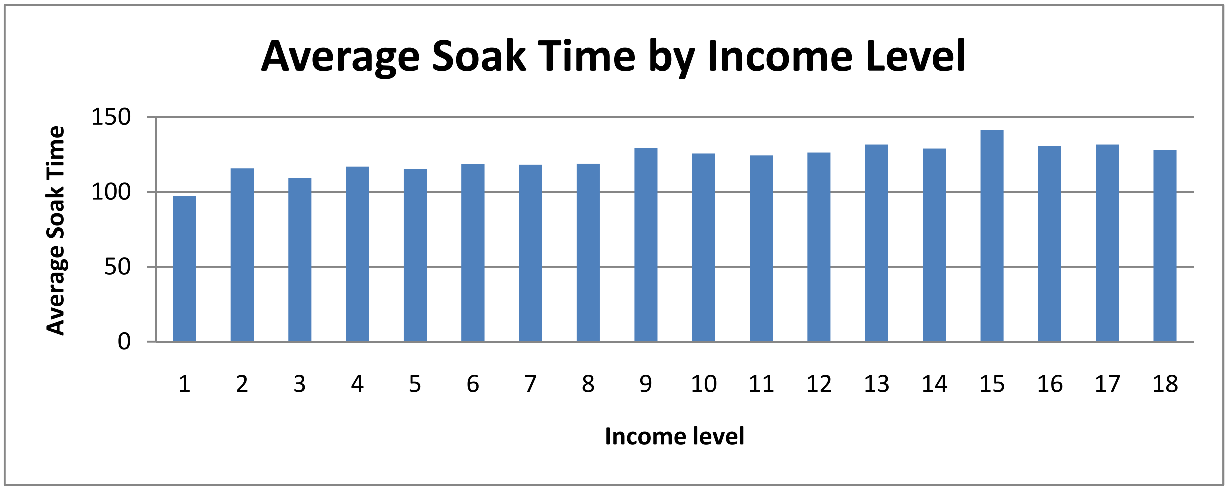
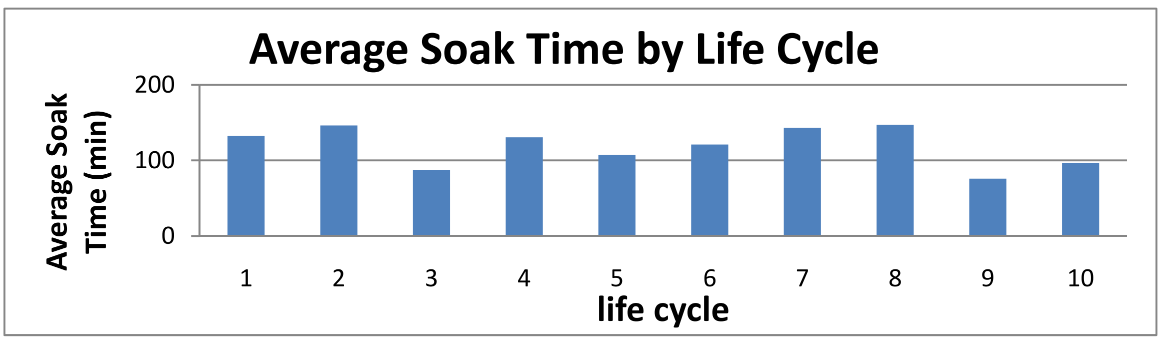


Figure 8. Distribution of Soak Time by Household Income Level



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 9. Distribution of Soak Time by Household Life Cycle

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

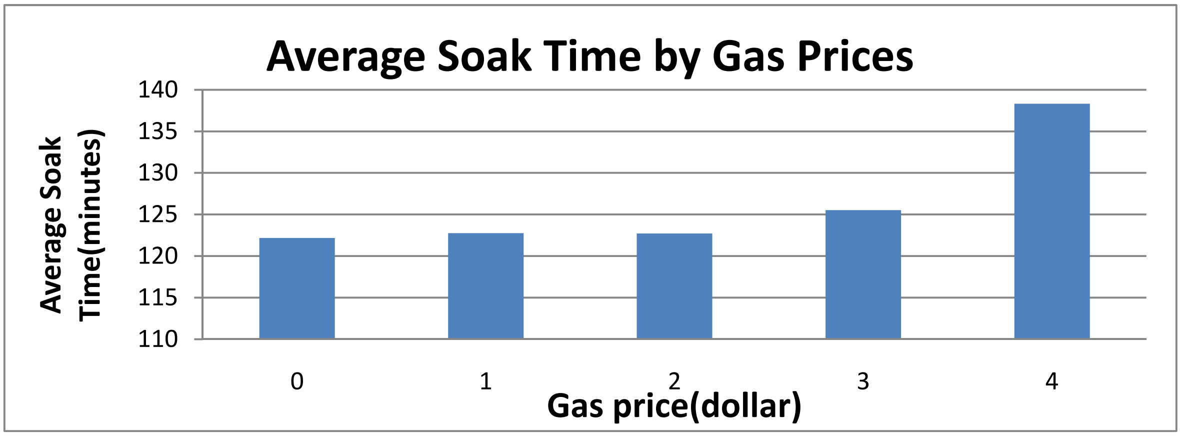


Figure 10. Distribution of Soak Time by gas price

One contribution in our research is the introduction of the gas price in soak time analysis. With the highest gas price, the soak time duration is longest, which is reasonable. It means people will adjust their several short trips into a long one, so they will have longer soak time duration than before. When the gas price changes less than 3 dollars, the soak time duration will change little because people pay little attention about the fewer gas prices changes.

5. Modeling of Soak Times

5.1 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. The logistic regression model results for first starts versus non-first starts are provided in Table 2. The base category used is non-first starts. Thus, a positive coefficient on the 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 differentiate the morning trips and weekend trips. The model results show that trip starts that occur 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, which means, everything else being equal, trips from home are most likely to be first starts. A comparison of the magnitudes of coefficients across the activity-purpose categories provides additional information 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 are not so likely is because quite a lot of work trips are observed to occur in the afternoon. We thereafter 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 2. Logistic Regression Model for First Starts

Variable	coefficient	P>z
Constant	-8.47447	
Soak Time Duration	0.0114227	0
Time of the Trip		
Morning Travel	3.168719	0
Weekend Travel	-0.0020471	0.987
Weekend Morning	-0.1559254	0.351
Activity purpose prior to trip start(Non-home purpose is base)		
Home-based Work(HBW)	-0.1453313	0.307
Home-based Shopping(HBS)	1.473298	0
Home-based Recreation(HBR)	1.841793	0
Home-based Other(HBO)	1.017323	0
HBW Morning Travel	1.422271	0
HBR weekend Travel	-0.5502737	0.038
Number of Observations	405536	
log likelihood	-30244455	
Rho-Squared	0.8827	

5.2 Model for Soak Time Duration for Non-First Start

Regression would be the normally used model to investigate the 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 soak duration, we could see clearly that the 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 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 we 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 soak time duration model for non-first starts are presented in Table 3. The dependent variable in the model is the soak time length. 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 the day variables indicate that the 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 soak time for the reason that people make more trips during weekends. The time of 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 soak period.

Table 3. Soak Time Duration Model for non-first Starts

Variable	coefficient	P
Constant	-240.2247	0
Time of 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	5.301924	0
Weekend PM Peak Travel	2.182414	0.253
Activity purpose prior to the soaking period(other purpose is base)		
Work and School	7.519454	0
Work related	-1.239013	0.488
Recreation, medical/dental	3.766162	0.001
Return Home	7.748846	0
Shopping	-1.658872	0.069
Serve Passenger	-6.199262	0
Lunch and dinner	-0.093967	0.931
Activity purpose of the non-first start trip(other purpose is base)		
Work and School	-3.229916	0.017
Work related	6.124929	0.04
Recreation, medical/dental	1.538946	0.224
Return Home	1.762309	0.061
Shopping	1.868356	0.081
Serve Passenger	1.676114	0.366
Lunch and dinner	1.916766	0.139
Interaction variables(time of day and purpose of the prior trip)		
Early Morning return home	-130.3941	0
Morning go to school/work	-7.659897	0.002
Wkend medical and recreation	0.1327257	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	-2.121943	0.232
Evening return home	-10.75231	0.065
Night go to work	4.162023	0.921
Weekend Travel	5.301924	0
Weekend PM Peak Travel	2.182414	0.253
Fuel variables		
gas price	2.022986	0.039
eiadmpg	0.1944711	0.108
Gas price*mpg	-0.0682226	0.122
Number of Observations	306446	
Rho-squared Stat	0.6	

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 soaking period after going to work or school. Thus we can observe the relatively higher coefficient for the work trips in the morning. We could 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. Thus they would have really long soak duration before going to work. We introduce gas price as an independent variable, and study the relation between the soak time duration and gas price. The coefficient of the gas price is positive; it means with the increase of the gas prices, the soak time duration will be much longer. It is highly possible that people will change their different short trip to a long trip to decrease the travel time, which definitely produce longer soak time. The interaction gas prices and mpg variables shows that even if the gas price increases, people will more likely choose the high mpg vehicle as the travel tool.

Same with the non-first start model, the first start model's results for time of the day variables indicate that the 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 will have a long soak time after they finish the whole day activity. And the return home purpose has the largest coefficient because people will have a long time for rest.

The negative coefficient of the gas price here means when the gas prices increase, the soak time duration will decrease. It is highly possible that people will change their departure time in the early morning to reduce the traffic congestion. There is another highly possibility that the season impacts the travel time then bring the results for the soak time reduction with the increased gas prices.

6. Conclusions and Future Research

From all the analysis and modeling we come up with the conclusions. Based on the distribution test analysis, the authors have observed that a number of attributes have great impact on soak time distribution, such as trip purpose, time-of-day, vehicle type, life cycle, and so forth. This helped us determine the variables we want to specify in the first/non-first starts model and duration model. 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, and thus it shows the necessity to treat these two types of trips independently in the soak duration model.

The soak-time duration has been modeled by this study 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. Also, we introduce the gas prices into all trips model; it 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 soak time duration.

References

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

Koupal, J., M. Cumberworth, et al. (2002). Draft Design and Implementation Plan for EPA's Multi-Scale Motor Vehicle and Equipment Emission System (MOVES). Washington D.C., EPA420-P-02-006, US Environmental Protection Agency, Office of Transportation and Air Quality.

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.

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.

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.

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

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

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.

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.

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.