Travel Behavior and the Effects of Household Demographics and Lifestyles

Taha H. Rashidi (Corresponding Author)
Ph.D. Candidate
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St. Chicago, IL 60607
Phone: 312-996-0962
Fax: 312-996-2426
Email: *thosse2@uic.edu*

Abolfazl (Kouros) Mohammadian, Ph.D.
Associate Professor
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St. Chicago, IL 60607
Phone: 312-996-9840
Fax: 312-996-2426
Email: *kouros@uic.edu*

Yongping Zhang, Ph.D.
Wilbur Smith Associates
1111 Burlington Ave
Lisle, Illinois 60532
Phone: 630-434-8111
Fax: 630-434-8163
Email: *zypcliff@gmail.com*

Total Words: 6,567 words

Abstract

Household and individual demographics, attributes and dynamics have significant effects on their travel behavior and the overall performance of the transportation system. This study attempts to study the effects of household demographic on the travel attributes of the household members of several homogeneous lifestyle clusters. Using the National Household Travel Survey (NHTS) 2001 data, more than twenty travel attributes including number of auto trips, trips per tour, transit usage and average commute distance are analyzed. To investigate the impact of changing demographics on household and individual level travel attributes, the best fitted distributions for a large set of travel attributes are introduced. Then the study provides a detailed comparison of the resulted distributions across different lifestyles and demographics.

Introduction

Travel demand models are used to predict future demand and its impact on the transportation system. Such models can provide accurate prediction results only if various dynamics and demographics are included in the model. However, practically, not many variables are included in the model, due to various reasons including data limitations as well as estimation and computational complexities. Recent developments in the area of travel data transferability have introduced robust procedures that are capable of considering many explanatory variables in an uncomplicated framework. Although, data transferability frameworks mainly replicate trip generation and mode choice steps of traditional four-step modeling, data transferability models are not as aggregate as the traditional four-step models are. Instead, highly disaggregate datasets are utilized for developing the models and estimating the coefficients. However, unlike activity-based models, data transferability models simplify the travel demand modeling procedure.

Therefore, from the methodological perspective, while these new approaches for forecasting travel attributes can be highly disaggregate, they are very simple and uncomplicated. Additionally, the data transferability framework is capable of considering a range of explanatory variables.

In this study, the authors aim to introduce a group of household clusters representing lifestyles in which members have similar travel attributes and characteristics. Effects of the household demographic variation between each pair of lifestyles on their travel attributes are analyzed in this study. In other words, effects of changes in household size, income, job type and many other socio-economic and demographic attributes on the household travel attributes like household number of trips per day are studied. The study is built upon an earlier research (Zhang and Mohammadian, 2008a) in which the data transferability framework considers eleven household lifestyle clusters. Given these clusters, the first preliminary but very significant analysis is to study the effects of the household demographics and life style on the cluster members' travel attributes.

More than twenty travel attributes are chosen to be studied. The study is performed using National Household Travel Survey (NHTS) 2001. Travel attributes such as number of auto trips, trips per tour, transit usage and average commute distance and many other variables are included among the set of the dependent variables.

The aforementioned clusters are identified and determined by using more than thirty explanatory variables. These variables compose of the NHTS 2001 household and individual socio-demographic attributes, as well as, other variables related to land-use, built environment, transportation system, and congestion related variables obtained from an extensive GIS analysis. Assigning cluster membership to all households in the NHTS 2001 dataset, the best fitted probability density functions for various travel attributes are fitted for each cluster. To accomplish this task more than forty distributions are tested for each travel attribute while considering the weight factors to expand the NHTS data to the entire US population. While in the transportation field, typically, normal distribution is mainly assumed and utilized as the primary distribution for travel attributes, this study shows that none of the best fitted distributions to the travel attributes are normal.

This study investigates the impact of changing demographics on household and individual level travel attributes. To achieve this goal, the study introduces the best fitted distributions for a large set of travel attributes. Nonetheless, results of distribution fitting procedure are usually rejected based on the Goodness-of-fit statistics due to the large sample sizes. Therefore, specific statistical analysis and data cleaning and preparation should be applied to find the acceptable distributions.

Then the study provides a detailed comparison of the resulted distributions across different lifestyles and demographics. In other words, different distribution types across the clusters are resulted because of the heterogeneity in population and variation in demographics and between the clusters attributes. However, determining the factors that provide the distribution differences requires further research which is performed in this study. For example, what is the primary reason why the senior citizens have lower trips per household? Why does the number of trips per person follow similar distributions across all clusters? Moreover, which one of the two mentioned travel variables (trip per household and trip per person) are more appropriate to be considered in transportation forecasting models? This study aims to provide answers to such questions.

Finally, since the results of this study are based on the NHTS 2001, therefore, the fitted distributions can be used for synthesizing travel attributes in other contexts. Where a local data source is available, the estimated distribution coefficients can be updated for the local areas using Bayes Theorem, so, the need for conducting a costly survey to collect travel attribute information can be eliminated.

Background and Methodology

Travel attributes are typically assumed to follow a normal distribution and the mean value and the standard deviation are the only parameters that are estimated while distribution of travel attributes are considered. However, the normality assumption for many travel attributes has been rejected in the literature (Zhang and Mohammadian, 2008a). Furthermore, it has been shown, particularly in activity-based micro-simulation models, that such distributions can be used at the disaggregate level to generate activities, trips, tours, modes, and other variables. Typically, a few explanatory variables are included in a linear or nonlinear travel forecasting model, like a trip generation model. However, where the travel attribute distributions are available across several homogeneous clusters, a larger set of explanatory variables can be included in the model. One can assume that the methodology behind clustering household into homogenous groups and finding the best fitted distribution to each travel attribute of each one of these clusters is similar to a cross- classification model where in each cell of the classification table there is a distribution instead of an average value. Such a simplistic approach can be used at a highly disaggregate level while the methodology behind it is not as complicated as the state-of-the-are micro-simulation models.

The National Household Travel Survey (NHTS) 2001 data, FHWA version 4 dataset 2005, which is a national inventory of household short- and long-term trips, is utilized in this study as the main source for developing the clusters and distributions. The data set includes 69,817 households with 642,292 trips that are representative of the general travel attributes of the nation in 2001. Given the large sample size and coverage of the NHTS, it has been shown that parameters derived from NHTS are transferable (Mohammadian and Zhang, 2007) and the quality of the transferred parameters can be significantly enhanced if updated using local data (Zhang and Mohammadian, 2008b).

Following a clustering procedure on NHTS, it is assumed that travel attributes are homogenous within these clusters and then various probability density functions (pdf) can be fitted to different travel attributes. One of the earliest attempts to fit distributions other than normal distribution to travel parameters was the study by Giannopoulos in 1977 where he found that Gamma family distributions provide acceptable distributions to trip length. Lau in 1999 also found negative exponential distribution to be a good fit for Trip Length. Number of trips per person and number of auto trips per person were the subject of another study in which gamma distribution was found to become the best distributions (Zhang and Mohammadian, 2008a). Nonetheless, many studies have assumed that the majority of travel attributes like VMT, and trip length follow a Gaussian distribution. Normal distribution is not only commonly used for travel attributes but it is also the most common distribution in the econometric models like regression. The coefficients of these models are considered to have a normal distribution based on central limit theorem (Hogg and Ledolter, 1992). The error term also assumed to have a normal distribution. However, the application of the central limit theorem is limited to several assumptions and should be verified in each case. Another reason for the great popularity of normal distribution is that commercial statistic and econometrics packages typically develop their results based on the assumption in which the coefficients are normally distributed.

Nonetheless, examining the distribution of error term and the heterogeneity term of different econometrics models have been a subject of research interest for many researchers. For instance, extreme value family distributions are commonly considered for the error term of the family of LOGIT models (McFadden, 1973) and gamma distribution provides good approximation to the unobserved heterogeneity in the hazard –based models (Bhat, 1996). It should be noted that there are several problems that should be considered when using normal distribution; for example, unbounded (sign problem), long tails and symmetrical shape (Hess *et al.*, 2006). Lognormal transformation can usually solve the first and the last problems, and the long tails problem can be solved by using the Triangular distribution (Hess *et al.*, 2006). Nevertheless, Gamma and Poisson distribution can solve each of these problems more effectively, as well as, offer the advantages of positive values and positive skew (Clark and Thayer, 2004).

In this study, more than forty distribution functions are tested for twenty four different travel attributes across households' lifestyles represented by eleven homogeneous clusters. Travel attributes are extracted from NHTS 2001 and are clustered based on the household demographic and socio-economic attributes, characteristics of the household residence, congestion level, and land use and built environment variable. Detail explanation for the clustering procedure can be found in a paper by Mohammadian and Zhang, 2007. In short, an array of more than 30 variables defines household clustering membership. Using a factor analysis procedure these variables are grouped into several independent factors that are later used in an artificial neural network model to assign NHTS 2001 households into eleven homogenous clusters. It has been shown that factorizing variables into smaller groups and clustering households into further homogeneous clusters can reduce the modeling error (Stopher 16*et al*. 2003 and Stopher *et al*. 2004). Once households are assigned to appropriate clusters, the best fitted probability density functions can be derived for each cluster and for various travel attributes.

Intuitively, it can easily be shown that household travel attributes are highly affected by household dynamics and demographic attributes. Many researchers considered household dynamics and demographic attributes in their models of household travel attributes forecasting. Ma and Goulias in 1997 clustered individuals and households into several clusters and studied their daily activity and travel patterns based on individuals' demographics and other dynamics using the Puget Sound Transportation Panel. Daily pattern of household travel and activity scheduling in the Toronto Area was studied by Miller and Roorda in 2003, in which they considered household's attributes and compositions types. There are many other studies and papers in this field in which travel attributes are modeled depending on household attributes; however, to the best of our knowledge, none of these studies have considered as many dependent variables as this study does. Moreover, this study presents a comprehensive descriptive analysis on the effects of numerous explanatory variables on various travel attributes that are categorized into household- and person-level.

Dependent variables and Cluster definition

Clustering households into homogenous groups in which households are assumed to have similar travel attributes can improve the quality of the fitted distributions. This study benefits from the clustering approach presented by Mohammadian and Zhang, 2008b, where over thirty explanatory variables including household residential location attributes, household socio-economic attributes, land-use and transportation system characteristics, and household demographics were utilized to identify eleven homogeneous clusters and they are listed in the next following lines (Mohammadian and Zhang, 2007).

1. **Rich and Smart:** This cluster represents middle-aged families with professional or managerial white collar jobs. They usually have graduate degrees and earn high incomes. The majority live in suburbs or towns. The ethnicity majority of this group tends Caucasian but there is Asian ethnicity as well.
2. **Young Achievers:** This group comprises of young couples without children or pre-school children. They tend to have college degrees and primarily have white collar jobs in sales, service, technical, and professional with mid-range income. There are higher percentages living in the suburbs or rural areas.
3. **Kids-centered Families:** These are middle-aged and working class families with pre-school and school-age children. They usually have a college education and earn mid-range to high level income. They are primarily Caucasian and live in suburbs or towns.
4. **Rural Blues:** The cluster includes working class, middle-aged families with pre-school and school-age children. They are mainly high school graduates in blue collar jobs (farming, manufacturing, etc) and earn low- to mid-range income. Mostly Caucasian and live in rural areas or small towns.
5. **Working Mixing Pot:** They tend to be working-class Caucasian, African-American, Asian, or Hispanic single adults or couples with college or high school education and low- to mid-range income. The majority live in suburban or rural area, but some in urban areas.
6. **Mainstream Families:** The cluster comprises of mid-scale, upper middle age, Caucasian, working-class couples or families with older children. They usually have a college or high school education, earn mid- to high-level income and live in suburb or rural areas.
7. **Senior Couples:** These are senior couples; primarily still working and some are retired. The majority of the group is Caucasian but it includes some African-American, Asian, or Native American. They live primarily in suburbs or rural areas.
8. **Sustaining Minority Families:** The cluster represents low-income, middle-aged, working-class families. They are mainly Hispanic or African-American but there are also some Asian and Caucasian. The majority have not finished high school. They tend to have service, sales, manufacturing, farming, or construction jobs.

9. **Forever Youngs:** These are Caucasian senior couples. Most of them are retired and empty-nested but some have sales, service, or managerial jobs and earn low- to mid-range income.
10. **Traditional Seniors:** They mainly comprise of retired single individuals. A number of them are retired couples with low income. The majority are Caucasian but some African-American, Asian, or Native American.
11. **Neo Urbans:** These are small families/couples or single individuals living in dense urban areas. They typically have a college education and low- to mid-range income from sales, service, or professional jobs. Their dominant race is Caucasian but a significant number are African-American, Asian, and Hispanic.

The dataset that was used in this study includes 26,038 households of the national sample of the NHTS 2001 excluding add-on observations. These households were grouped into the abovementioned eleven clusters. Then twenty four key travel attributes of households were selected for further analysis in this study. The list of travel attributes is selected so that they satisfy many of the transportation modeling data needs including trip generation and mode choice, which are two main elements of the traditional four-step modeling framework. The dependent variables are highly disaggregated and at the household level so they can be also utilized by state-of-the-are activity-based models. A portion of these dependent variables provide information on number of household trips grouped by trip purpose. The household number of auto, transit and non-motorized trips are also modeled in this study. Trip length and commute distance are other travel attributes that can be found useful in destination choice models. Moreover, trip length, commute distance and household annual Vehicle Mile Traveled (VMT) are considered that are commonly used in emission models. The household number of tours and number of trips per tour that are included in the analysis of this study can be effectively applied in activity-based models. The best significant distributions for all of these dependent variables are identified at the household and person levels and are presented in this paper.

Distributions and Goodness-of-fit

More than forty probability density functions are considered in curve fitting exercise of this study, among which eighteen distributions were found to be statistically significant for at least one of the travel attributes. Furthermore, it became clear that there are eight distributions that are selected more than any other observations. These are presented in Table 1 along with the times that they have been selected as the best distribution in 264 cases (11 clusters × 24 travels attributes).

Table 1 Eight of the distributions which are most selected as the best fitted distribution

Distribution Name	Gamma	Lognormal	Weibull	Generalized Extreme Value	Logistic	Log-Logistic	Johnson SB	Gumbel Max
Times Selected	9	10	19	21	23	35	37	78

The distribution fitting process is performed by employing a statistical and mathematical package called EasyFit to test several probability density functions and selects the best fitted distributional forms. The software tool has also the capability of considering the observation's weight factors. It was shown that the inclusion of weight factors considerably affects the final results. In order to evaluate the quality of fitting exercise, two goodness-of-fit measures can be considered including the Anderson-Darling test (AD) (Anderson and Darling, 1952), and the Kolmogorov-Smirnov test (KS) (Chakravarti *et al*., 1967 and Eadie *et al.*, 1971). The AD test is one of the most dominant statistics for finding most departures from normality. On the other hand, the KS test is utilized to verify whether two underlying one-dimensional probability distributions vary, or whether an underlying probability distribution differs from a hypothesized distribution. Additionally, the KS test is just used on databases with finite sample sizes. Since in this study, neither the underlying data nor the distributions is necessarily normal therefore the KS test should be considered as the critical statistic for selecting the best fitted distribution. It worth noting that Kolmogorov-Smirnov test is highly sensitive to the number of observations and it has a higher probability for rejecting the null hypothesis for large samples. The spurious goodness-of-fit statistics which are formed by large samples used in a KS test are adjusted into meaningful values by grouping together the observations with similar dependent variable values. This action reduces the number of observations without changing the frequency of the observations and the weight factors of the grouped observations are summed up together. This simple method solves the problem of spurious KS statistics, which might result in rejecting distributions that are statistically significant. Nonetheless, either when the grouping method is used or in the absence of this method the coefficient estimation is unchanged.

In short, by considering the KS method as the dominant statistic for selecting the best distribution, 264 distributions are calibrated for eleven clusters across twenty four travel behavioral attributes. The results of the parameters of these distributions are presented in the results and analysis section.

Results and Analysis

In this section detailed numerical analysis for several household trip variables are presented. Additionally, graphical analysis of the overall household travel attributes grouped by different clusters is presented as well. Tables 2 and 3 show the detailed estimated parameters for each household and individual number of trips. Two travel attributes for eleven clusters collectively make twenty two distributions that are presented in these two tables. In this study twenty four tables like the ones presented here are developed for various travel attributes that were listed earlier. Although, due to page limitations, the entirety details of the distribution parameters is not presented in this paper, they are available through authors upon request for further analysis and application in other research studies.

Table 2 Distributions for Household daily trips for different clusters

Trip Per Household

Cluster	Distribution	Parameter								Kolmogorov-Smirnov		
										Statistic	Rank	P-value
1	Gumble Max	σ	6.6018	μ	11.034					0.05133	1	0.99873
2	Lognormal	σ	0.52087	μ	2.5866	γ	2.2317			0.044999	1	0.99979
3	Johnson SB	γ	8.7298	δ	2.9292	λ	578.51	ξ	-12.343	0.03182	1	1
4	Johnson SB	γ	3.8906	δ	1.9907	λ	151.89	ξ	-4.6891	0.0409	3	0.99998
5	Weibull (3P)	α	1.3931	β	5.1494	γ	0.93309			0.09406	1	0.9406
6	Log Logistic (3P)	α	3.3829	β	13.119	γ	-1.7378			0.04813	1	0.99956
7	Weibull (3P)	α	1.6092	β	9.3165	γ	0.80181			0.05772	2	0.99949
8	Fatigue Life	α	0.58117	β	12.568	γ	-2.3197			0.04627	1	0.99988
9	Johnson SB	γ	6.4402	δ	2.6073	λ	168.54	ξ	-4.9905	0.07025	1	0.99515
10	Gamma	α	2.8769	β	1.6014					0.12321	2	0.86937
11	Weibull (3P)	α	1.3231	β	6.1337	γ	0.922			0.08909	1	0.98214

It can be seen in the last column on the right hand side of these two tables that all of the fitted distributions cannot be rejected at the 0.01 confidence level. In other words, the difference between the fitted distribution and the data is not significant. One can observe that in Tables 2 and 3, sometimes distributions that are not ranked first are selected. This is due to the fact that some distributions like Phased Bi-Exponential, Phased Bi-Weibull and Wakeby are selected frequently as the high ranked distributions; however, these distributions have highly complicated formulations with five parameters or more. Due to their complicated formulations for a simulation application and their unpopularity these distribution were disregarded in this study. Nonetheless, the K-S statistic of the ignored distributions and the selected ones are very close to each other.

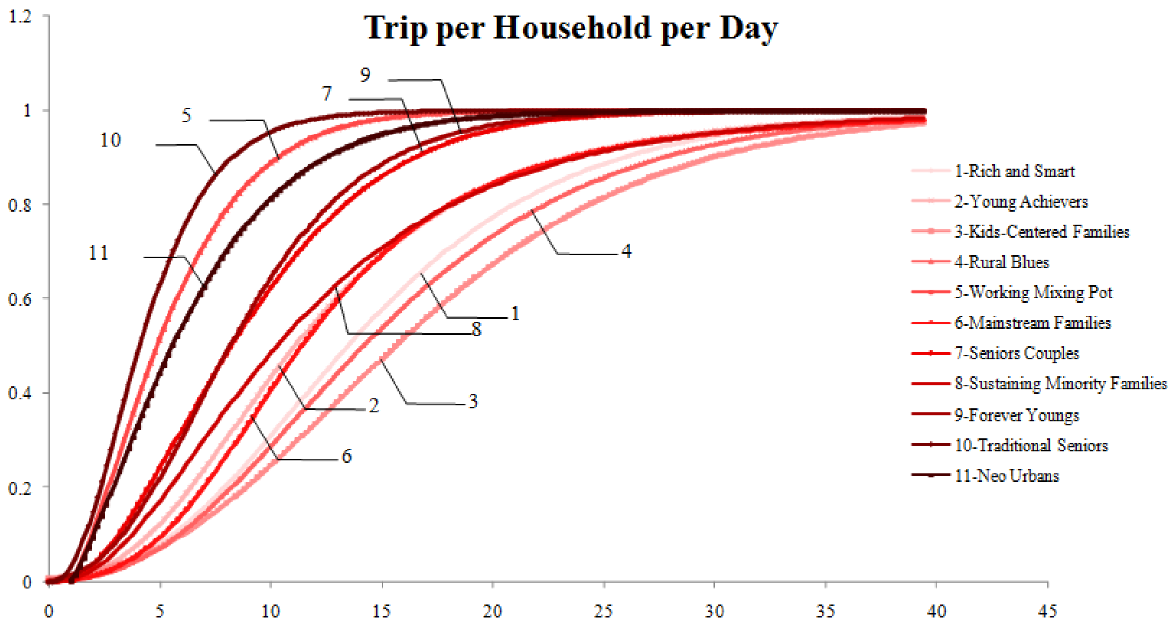
Table 3 Distributions for individual daily trips for different clusters along with the KS and AD statistics

Auto Trip Per Household

Cluster	Distribution	Parameter								KS		AD	
										Statistic	Rank	Statistic	Rank
1	Gumbel Max	σ	6.0736	μ	9.3896					0.0585	3	2.1751	1
2	Gen. Extreme Value	κ	0.11426	σ	5.225	μ	7.819			0.0456	1	3.6587	1
3	Johnson SB	γ	5.447	δ	2.4087	λ	242.99	ξ	-9.4669	0.0332	1	2.0607	1
4	Johnson SB	γ	3.526	δ	1.8525	λ	130.79	ξ	-4.7981	0.0472	1	2.8108	1
5	Gumbel Max	σ	2.6749	μ	3.4274					0.1006	2	34.025	1
6	Gamma	α	2.697	β	4.4286					0.0467	1	29.269	1
7	Gumbel Max	σ	3.9953	μ	6.1107					0.0686	2	8.2661	1
8	Gen. Extreme Value	κ	0.12606	σ	5.5539	μ	6.4228			0.051	1	4.5956	1
9	Johnson SU	γ	-7.7121	δ	3.1757	λ	2.4828	ξ	-6.5584	0.0802	1	8.9271	1
10	Gen. Extreme Value	κ	0.06954	σ	2.1208	μ	2.8633			0.1293	1	23.24	1
11	Gumbel Max	σ	3.0242	μ	1.6441					0.1787	1	17.582	1

The main part of the study involved descriptive analysis of the changes of dependent variables across the eleven clusters and evaluating the influence of clusters' characteristics and attributes on observed travel parameter changes. The first set of the dependent variables examined in this study include household number of trips and individual number of trips. The distributions of these two dependent variables are presented in Figure 1.

In Figure 1, daily household and individual number of trips are presented through their cumulative density functions across the eleven clusters. A basic comparison between the general trends of the both figures depicted in Figure 1 reveals that household number of trips strongly depends on household attributes and the cumulative distributions are significantly different among the clusters. However, it is shown that the number of person's trips do not vary significantly across eleven clusters. In the upper figure, the curve for cluster 3 stands beneath all of the other cumulative distribution functions which means probability of making more trips is higher for cluster 3. Households in cluster 3 are typically larger households both in terms of number of workers and number of people; therefore, it is not surprising that they make more trips than other clusters like single seniors. Cluster 4 stands to the left of cluster 3 and is very similar to it in terms of its socio-economic attributes. These are households living in suburbs rather than rural areas. Furthermore, it can be observed that there are two groups of curves (group 1: clusters 6, 2 and 8; group 2: clusters 7 and 9) that are close to each other even with some overlap. It can be noticed that while clusters 6, 2 and 8 may have similar distributions of household number of trips, they are very different in terms of their level of education, income, ethnicity and household size. Similarly, clusters 7 and 9 have very close distributional curves which implies that senior couples have a similar patterns of number of trips. It appears that employment, which is the main difference between these two groups of seniors, is not a significant factor on the number of trips they make. Finally, cluster 10 households have the least number of trips which can be rationalized by their small household size.



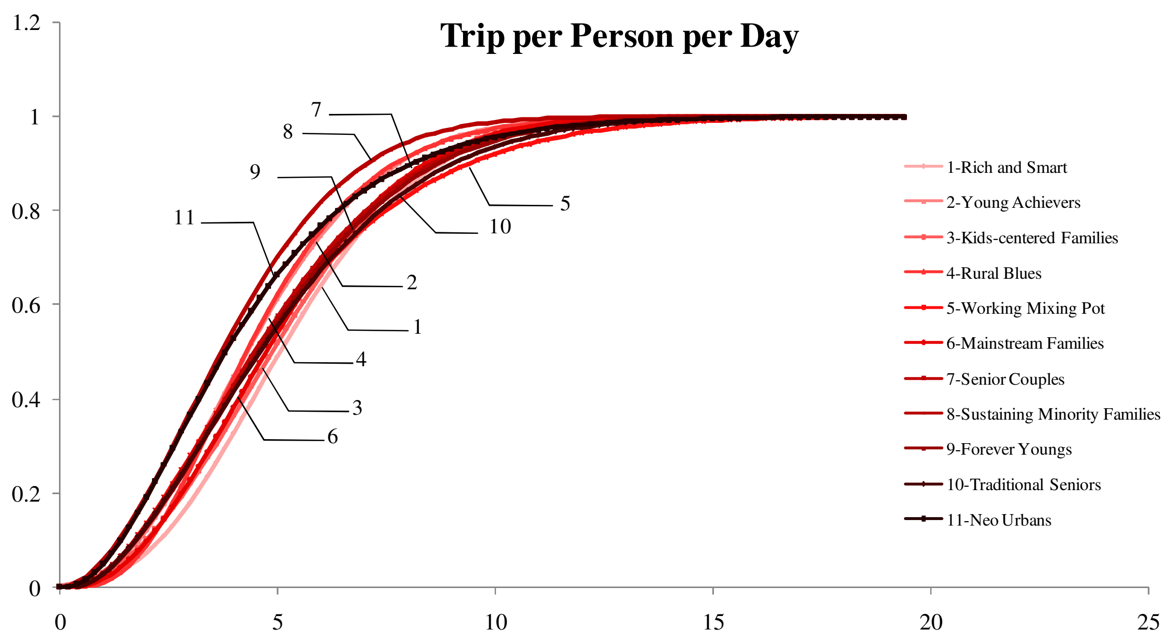


Figure 1 Number of trips at household and person levels for the eleven clusters

The next sets of dependant variables presented here are household and individual level number of auto trips. Figure 2 shows the results of fitted distributions to these two variables.

Like in Figure 1, the curves in Figure 2 show a similar number of auto trips trend per person across all clusters except for people living in urban areas where alternative modes are available and the driving mode can be more expensive. However, at the household level various clusters have different auto trip cumulative density functions. Cluster 3 again has the highest probability of having more auto trips and it is followed by clusters 4 and 1. Clusters 6, 2 and 8 are again very close to each other when number of auto trips is greater than 15, while probability of making auto trip for cluster 8 is decreased when number of trips is less than 15. Senior couples of clusters 9 and 7 both have similar auto trip patterns which again imply that employment is not a significant factor in household level trip generation for elderly people. Clusters that have the least number of auto trips include clusters 11, 5 and 10. Households of cluster 10 drive less and as shown later in this section, they may prefer non-motorized mode. Another interesting point which can be observed in the household level figure is that the curve for cluster 8 crosses the curves of clusters 7 and 9 at a number of auto trips close to 5. This implies that cluster 8 has higher probability of making more trips while the number of trips is higher than 5 compared to clusters 7 and 9, and vice versa.

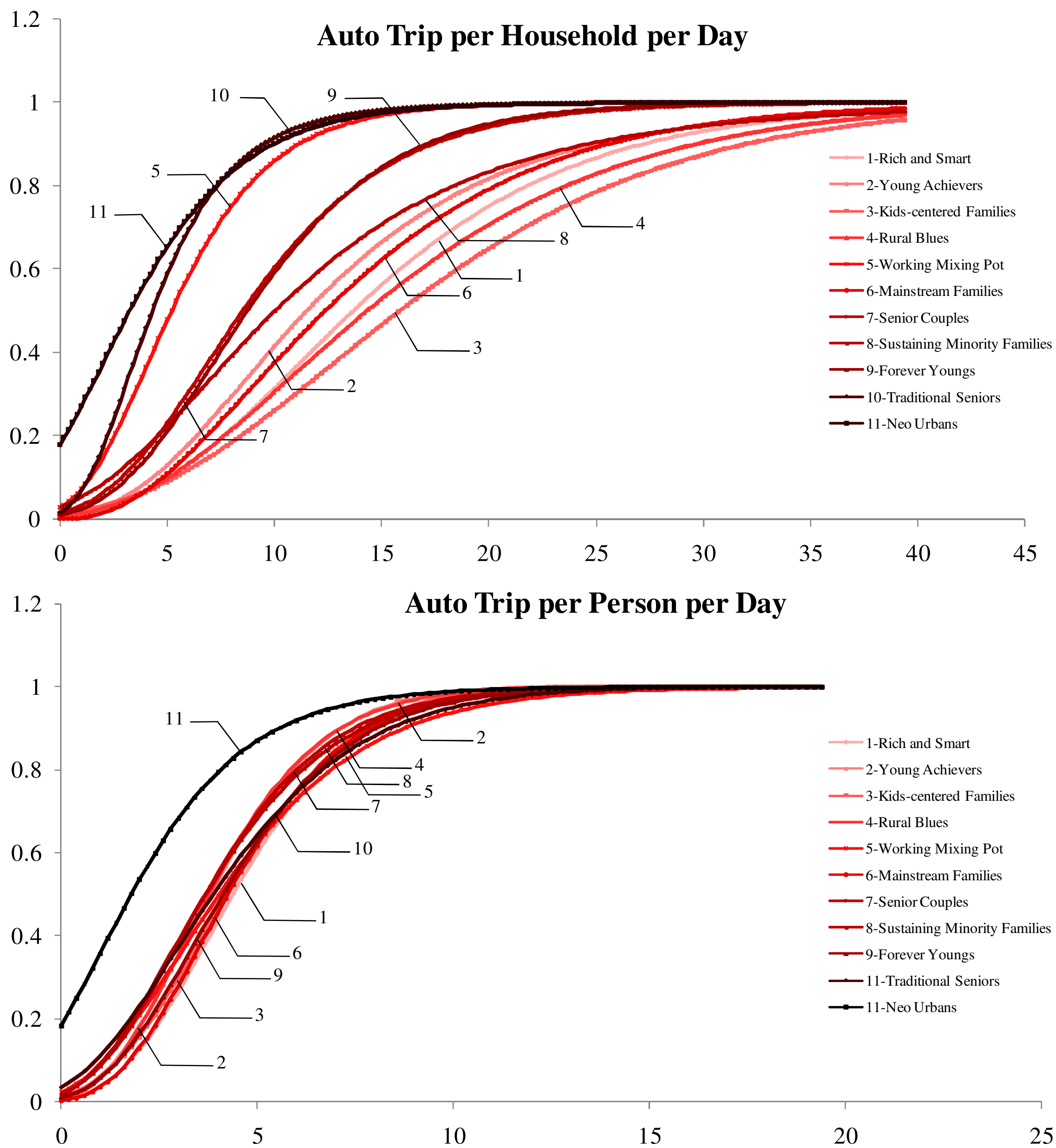
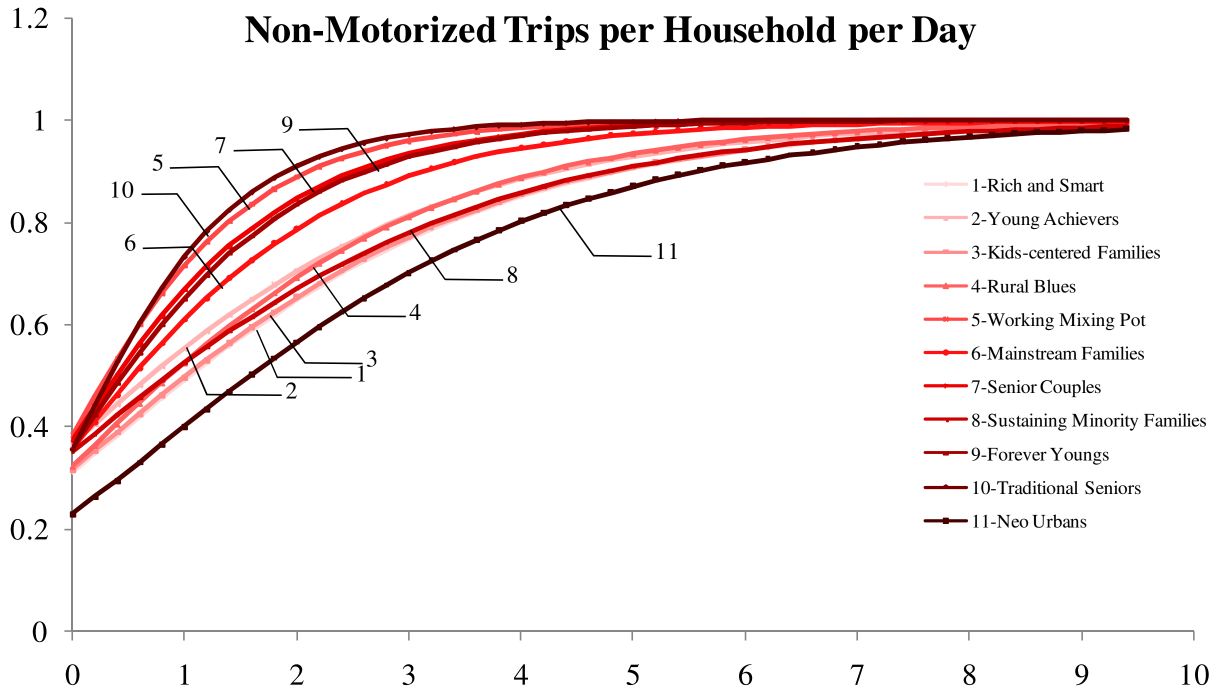


Figure 2 Number of auto trips at household and person levels for the eleven clusters

Figure 3 presents the cumulative density functions for non-motorized trips at household and person levels for the eleven clusters. At the person level, the individual members of the 11th cluster have the highest probability of making more non-motorized trips which is reasonable due to availability of alternative modes and the cost of driving in urban areas and also probably due to shorter trip distances. Among the other clusters, cluster 10 stands as the second cluster with the highest number of non-motorized trips after cluster 11 but with a great marginal distance. On the other side, cluster 6 has the lowest probability of making non-motorized trips which can be explained by the fact that households in cluster 6 have older children and largely live in rural or suburban areas where travel distances are relatively long. Interestingly, seniors grouped in cluster 10 make a great portion of their trips in a non-motorized mode which is consistent with what had been explained earlier for the auto trip results. More interestingly, households in cluster 10 are mainly single elderly people and the cumulative density function for this cluster is almost unchanged at household level and person level. However, the number of non-motorized trips for other clusters other than cluster 10 is highly affected by household size, therefore, it can be seen in the household level figure that cluster 10 has the least number of non-motorized trips at the household level. Consequently, one can conclude that household size and number of non-motorized trips are highly correlated.



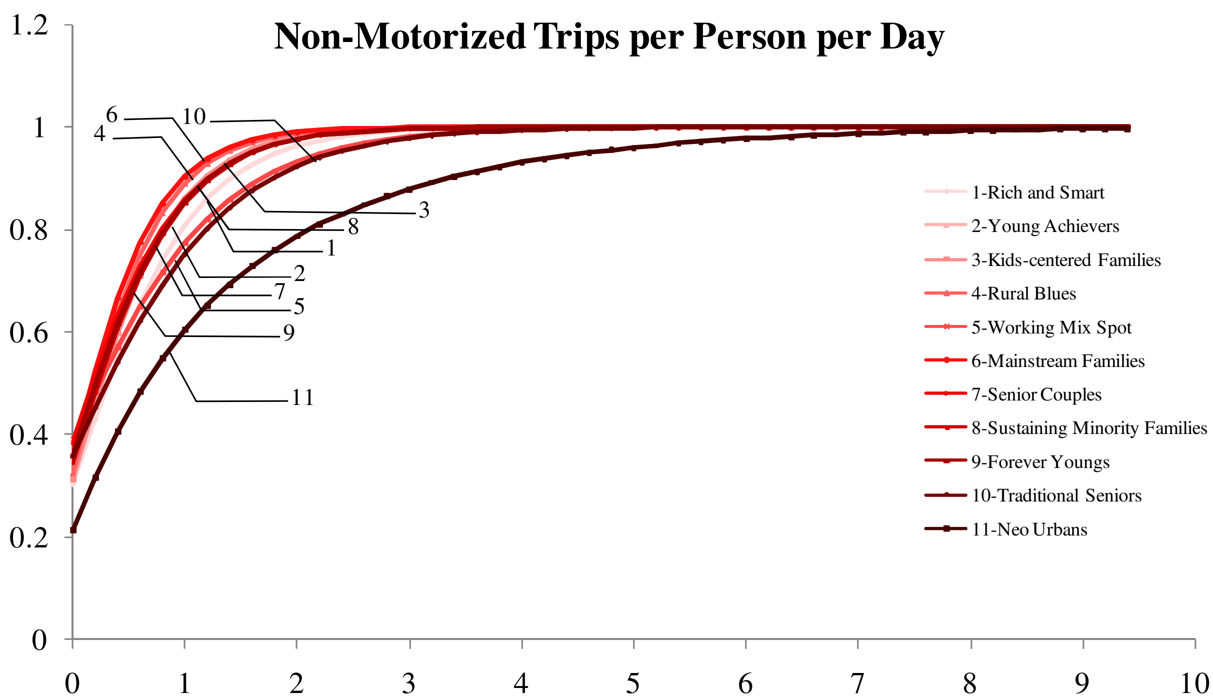


Figure 3 Number of non-motorized trips at household and person levels for the eleven clusters

The result of examining market shares of transit mode is presented in Figure 4. The general pattern of the graphs in Figures 4 is similar to the ones presented for non-motorized trips in Figure 3. The highest rate of transit usage is for people who live in dense urban areas. At the household level, all of the three clusters that represent seniors show less transit usage at the household level; however, at the individual level cluster 10 shows higher probability of having more transit usage probably because of their small household sizes. In other words, low income single Caucasian seniors use transit more often compared to other people who live in suburbs or rural areas. Cluster 8 stands at the second level of having the higher transit usage after people who live in urban areas. The majority of the households in cluster 8 are low income minorities who have barely finished high school.

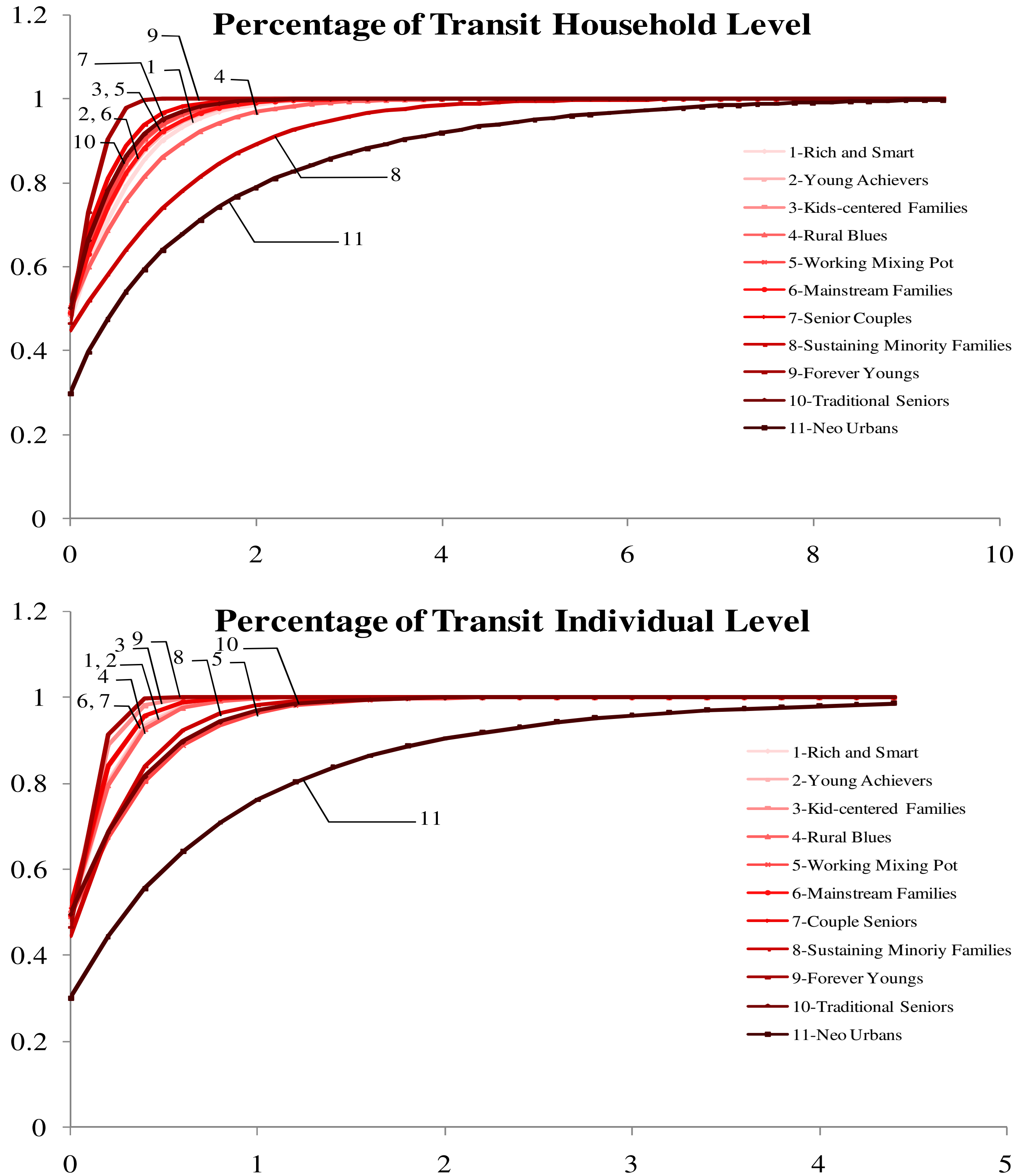


Figure 4 Number of transit trips at household and person levels for the eleven clusters

The next set of travel attribute variables are categorized by trip purpose into three groups: mandatory trips (e.g., work, school), maintenance trips (e.g., shopping, banking, visiting doctor) and discretionary trips (e.g., social and recreational activities, eating out). Figure 5 shows the cumulative density functions for these trip categories. As expected and confirmed by these figures, seniors have the most number of maintenance and discretionary trips while they have the least number of mandatory trips. Clusters 5 and 6 have the highest probability of making more mandatory trips where cluster 5 has even higher rate of mandatory trips when the average number of trips per person is greater than 1. In another view, the probability of making maintenance and discretionary trips is smaller for low income households in cluster 8. Households in cluster 8, in which most of the minorities are grouped, have mainly service, sales, manufacturing, farming and construction jobs.

For the mandatory trips, cluster 3 is beneath the rest of the clusters with clusters 1 and 4 directly above it. The main similarity between these three clusters is that all their corresponding households live in both suburbs and rural areas; therefore, they might have multiple destinations for their mandatory trips which increases their number of mandatory trips.

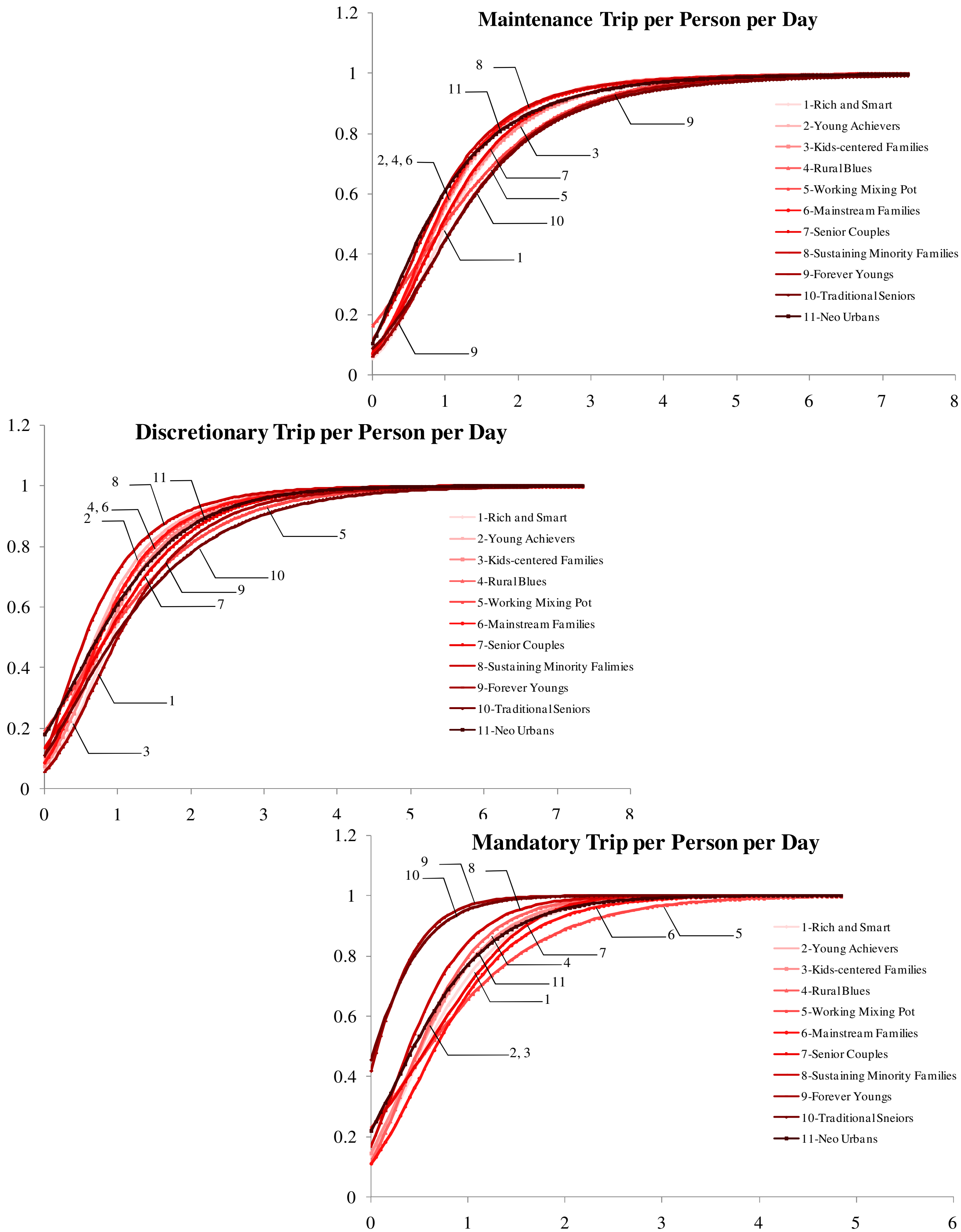


Figure 5 Number of mandatory, maintenance and discretionary trips at person level for the eleven clusters

Vehicle Miles Traveled (VMT) is the next variable for which the modeling results are presented in Figure 6. At the person level, clusters 5 and 11 are the ones with the highest probability of driving more miles. The common attribute between these two clusters is their small household size, income and education levels. The curves for clusetrs 7 and 10 ,which are both elderly people with different employment status, crosses each other at a point close to 5000 miles. It can be concluded from this figure that employed seniors tend to drive less than retired ones when the VMT is 5000 miles or more and vice versa.

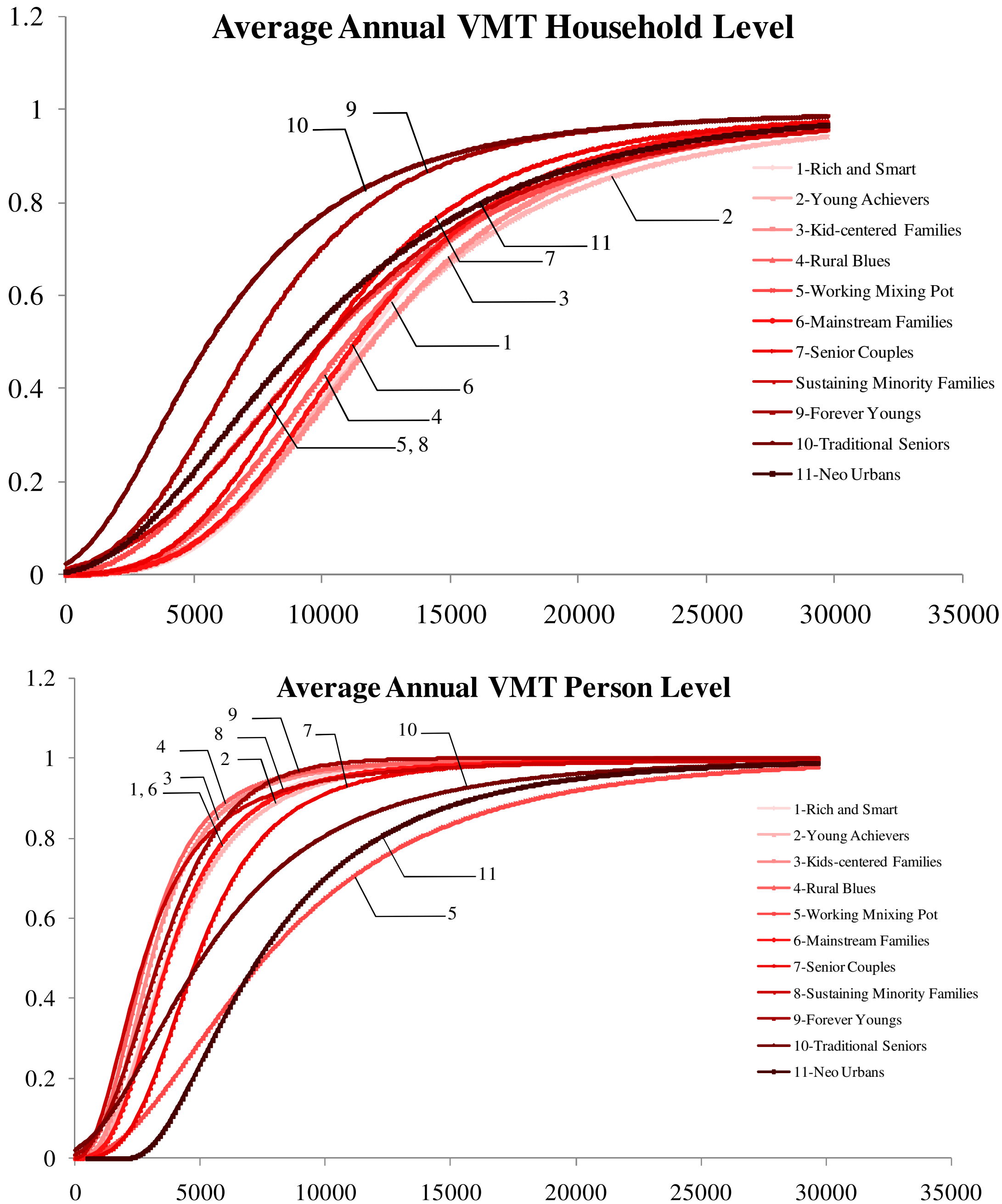


Figure 6 Average annual VMT at person and household levels for the eleven clusters

The results also suggest that at the household level, among all groups, the probability of driving more is the highest for clusters 1 and 2. The small households that can be categorized into clusters 9, 10 , and 11 have the lowest annual average VMT. It should be noted that cluster

11 at the household level with a VMT value greater than 15,000 has similar pattern to other clusters, but at the person level this group has the lowest VMT value. This questionable result might be resulted from the fact that reported VMT in the NHTS 2001 is not the actual odometer values but are the best estimates and therefore the approximation on the VMT values are not neccessirily accurate.

Another interesting dependent variable which is estimated in this study is the average trip length. The modeling results for this variable are presented in Figure 7.

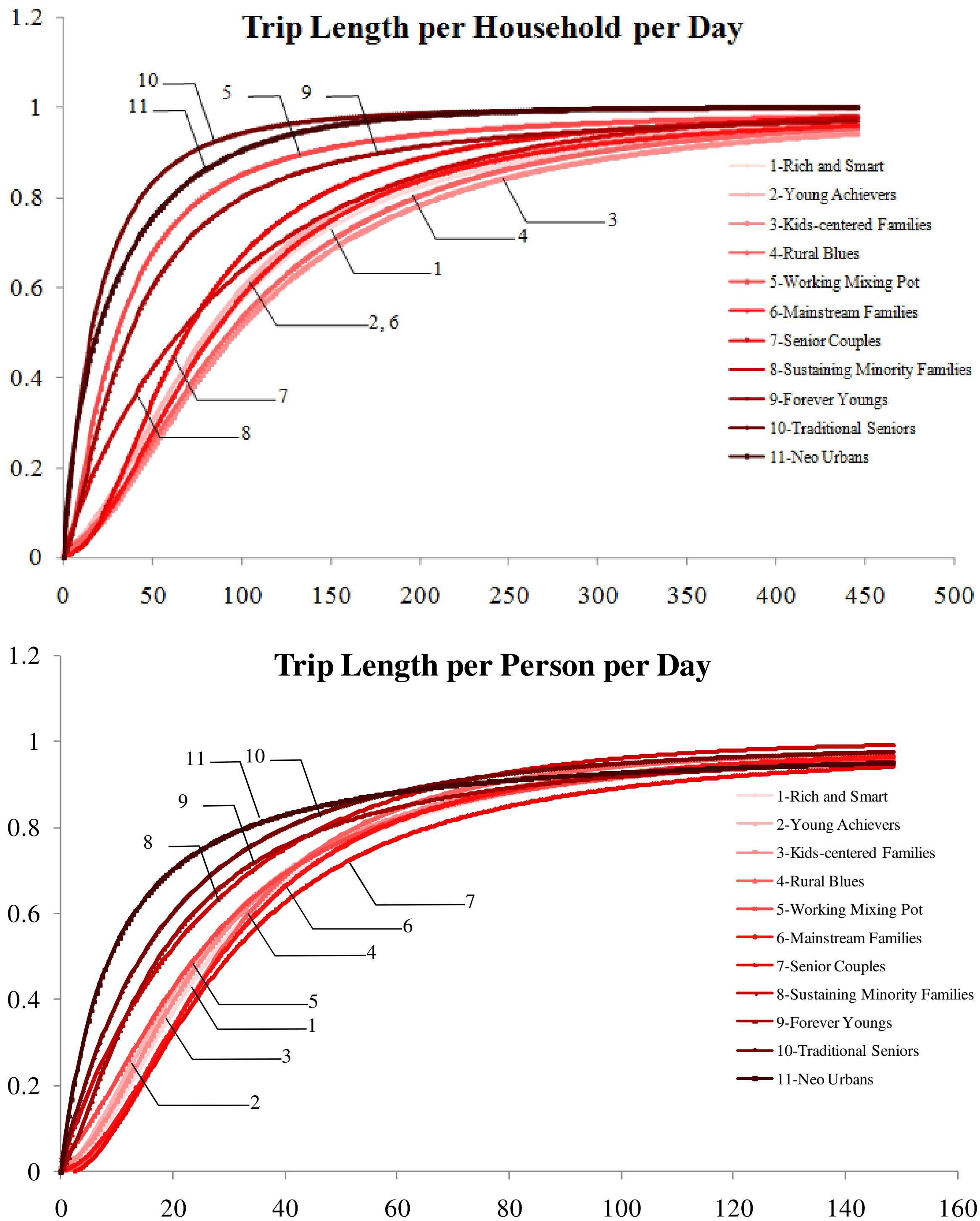


Figure 7 Average trip length in miles at person and household levels for eleven clusters

It can be shown that at the person level, families who live in dense urban areas often make shorter trips which is also consistent with the logical expected results. It appears that in addition to cluster 11, clusters 9 and 10 (seniors) prefer to make shorter trips. Nonetheless, families in cluster 7 which are suburban employed seniors with high incomes, have the highest probability of making longer trips. At the household level, similar patterns can be seen for clusters 9, 10 and 11 who prefer to select destinations which are closer to their origin. However, cluster 5, that mainly represents working singles or couples, stands close to these three clusters. Unlike the person level, at the household level clusters 2 and 3 are shown to have longer trips because they both live in suburbs and they both have school age children. They also have white-collar jobs.

Another dependent variable which has a close correlation with trip length is commute distance. Most likely, the household level average commute distance figures are quite similar to what was shown for trip length where clusters 5, 9, 10 and 11 were the ones with the shortest commute distances and clusters 3, 4 and 6 were the ones with the longest commute distance. Therefore, it can be concluded that households with similar household sizes have similar commute distance patterns, even though they have completely different socio-demographic characteristics. However, at the person level, clusters with many pre-school or school age children fall at the shortest commute distance position. These clusters include clusters 3 and 4 and their distributions are positioned above the rest of the clusters. On the other side, clusters 2, 6 and 7 which are mainly households without children seem to have the farthest work destination among all of the other clusters.

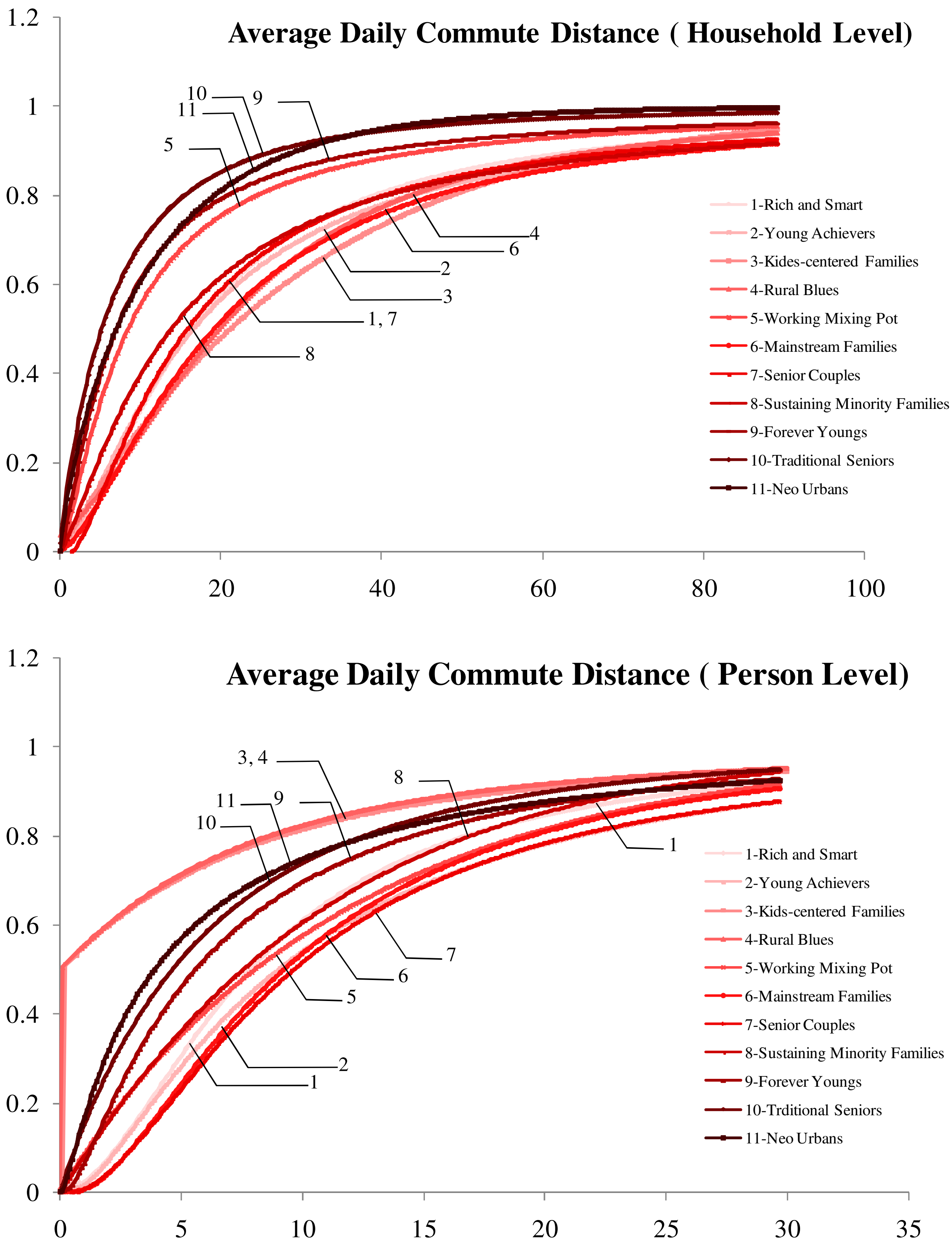
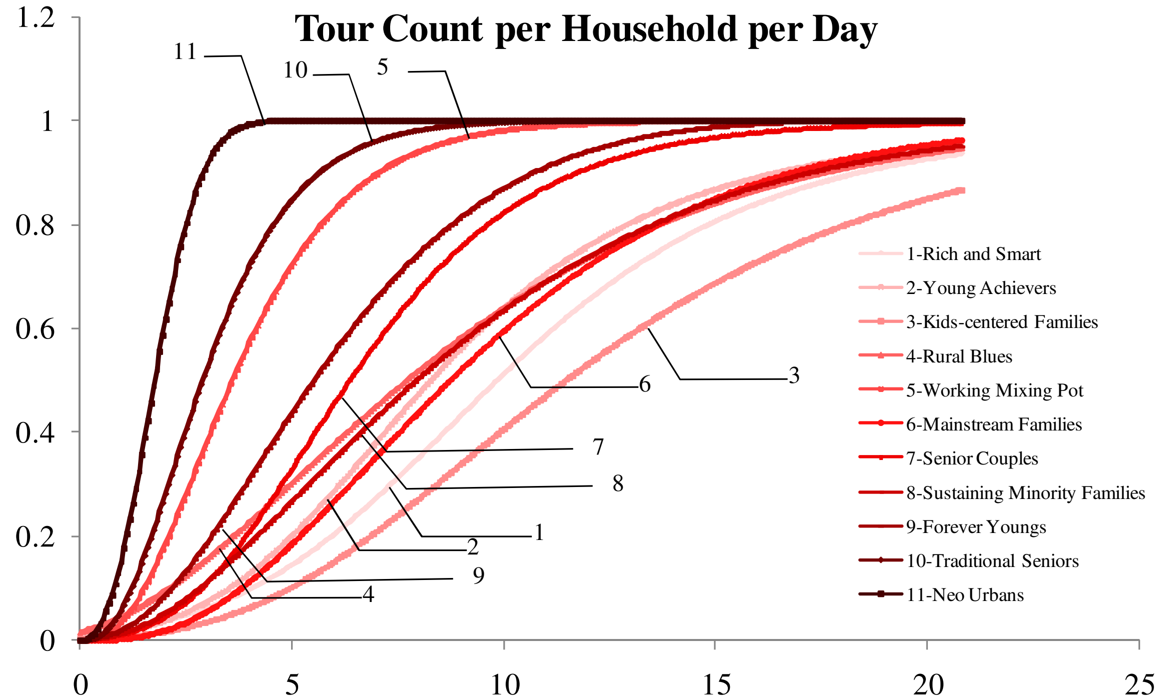


Figure 8 Average commute distance in miles at person and household levels for the eleven clusters

The last sets of travel attribute variables considered in this study refer to the number of tours and number of trips within each tour. Figure 9 shows the result for these two variables at person level.



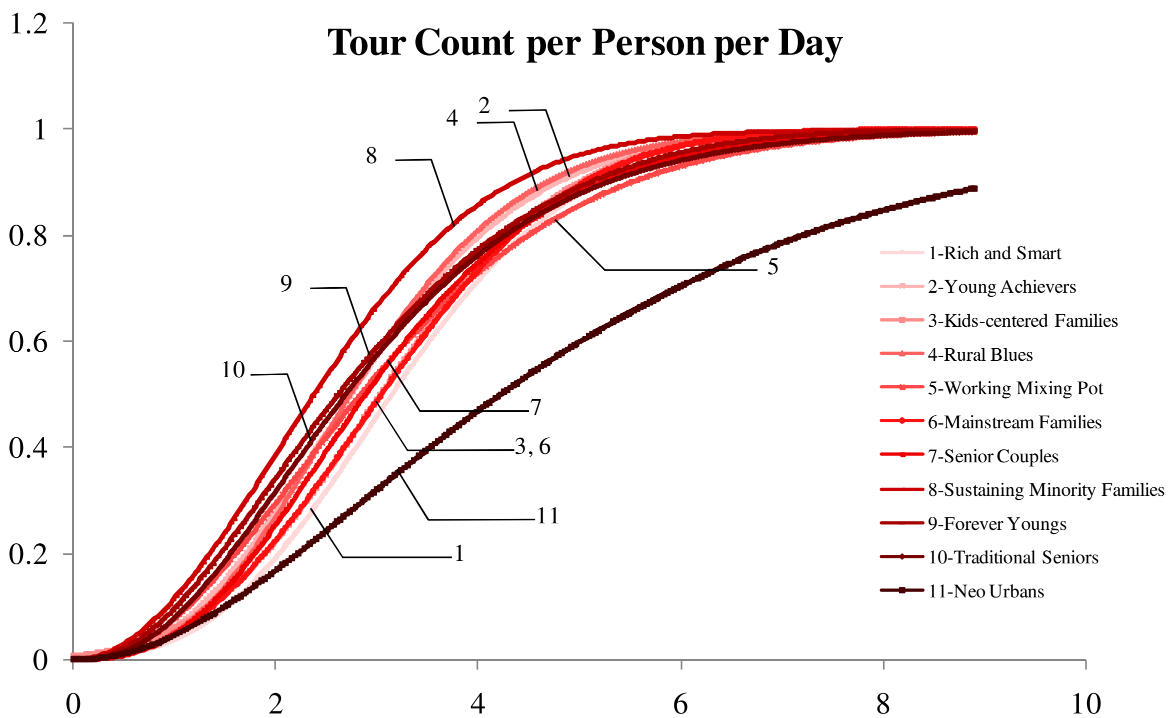


Figure 9 Number of tours at person and household levels for the eleven clusters

Families that live in dense urban areas tend to make more tours and the average number of trips within each tour for them is considerable. However, retired single seniors, cluster 10, have the greatest number of trips within each tour and minorities, clusters 5 and 8, have the least number of trips per tours. Cluster 8 which mainly composed of minorities stand above the other clusters in the tour count figure of Figure 9. Minority families also have the second lowest number of trips within each tour after cluster 5. Cluster 5 also includes families with low incomes and it also includes households from a variety of ethnicities.

The results of the analysis of fitting best distribution functions to many travel attributes that are presented here suggest that contrary to the common practice, most of the interactions of travel attributes with household socio-demographics, land-use and congestion characteristics do not follow a normal distribution.

Conclusion and Future Tasks

Traditionally, travel attributes were considered to follow a normal (Gaussian) distribution. Using the National Household Travel Survey (NHTS) 2001, several probability density functions are examined with acceptable goodness-of-fits. This study showed that the normality assumption can be questionable at least for the travel attributes tested here. The distribution tests were done for eleven clusters of homogeneous household types representing their lifestyles. Households were assigned to each cluster based on their socio-economic, demographic, built environment, residential location attributes, and characteristics of the transportation system in their home zone. For each cluster, twenty four household- and individual-level travel attributes were considered and the best fitted cumulative (probability) density function (pdf) was selected for each travel attribute. To do so, more than forty different pdfs were tested and validated.

Then, a detailed analysis of the effects of changes to household dynamics on their travel attributes were presented and discussed. The descriptive analysis showed that some of the travel attributes were similar at person level and do not change greatly among the different clusters. For instance, total number of trips at the person level had a similar tendency among all clusters; however, some other clusters have significantly different pdfs (e.g., percentage of transit usage and tour counts). Employment was found to be a significant explanatory variable for distinguishing different senior clusters. Employed seniors typically have different travel attributes than unemployed elderly people. Land-use variables such as whether the household is located in a suburb or dense urban areas can greatly affect its travel attributes such as commute distance, tour count, non-motorized trips and transit usage. Household size was found to be highly correlated with trip count and auto-trip count. It was observed in this study that the curves obtained for the reported Vehicle Miles Travel (VMT) in NHTS 2001 were not consistent with other variables like commute distance and trip length.

This study introduced a descriptive, useful and applicable methodology to examine the impacts of changing demographics on household- and individual-level travel attributes in which many explanatory variables can be included. Moreover, the simplicity of the approach makes it easier to capture more dependent variables. The results of the study can be also useful for transportation simulation applications where household travel attributes should be randomly drawn. Additionally, it should be noted that using the Bayesian updating approach the obtained probability density functions can be easily updated for every level of local area (Block Group, Census Tract, City, MSA, etc) for which a small sample data is available.

As a future work and alternative approach, one can use other clustering approaches like Latent Class Models to develop different and further detailed homogeneous clusters. The results of this study can also be used in a data transferability model where the simulated and synthetic results can be compared against other existing datasets.

References

1. Anderson, T. W. & Darling, D. A., 1952, Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Annals of Mathematical Statistics, 23,* 193–212.
2. Bhat, C. R. (1996). A hazard-based duration model of shopping activity with nonparametric baseline specification and non parametric control for unobserved heterogeneity, *Transportation Research*, *30B*, **3**, 189-207.
3. Chakravarti I.M., R.G. Laha, J. Roy, 1967, Handbook of Methods of Applied Statistics, vol. I, John Wiley and Sons.
4. Clark, D.R., Thayer , C.A., 2004, A Primer on the Exponential Family of Distributions, *2004 Call Paper Program on Generalized Linear Models*, Accessed July 20th, 2006.
5. Eadie, W.T., Drijard D., James F.E., Roos M. and Sadoulet B., 1971, Statistical Methods in Experimental Physics. Amsterdam: North-Holland, 269-271.
6. FHWA , 2005, National Household Travel Survey (NHTS), version 4 datasets, http://nhts.ornl.gov/2001/index.shtml, accessed in July 2005.
7. Giannopoulos, G. A., 1977, An Analysis of Trip Length and Land Use Patterns in the. *Transportation , 6*, 379 -392.
8. Hess, S., Axhausen , K.W., Polak , J.W., 2006, Distributional Assumptions in Mixed Logit Models, *85th Annual Meeting of Transportation Research Board*, Washington D.C.
9. Hogg, R. V., & Ledolter, J., 1992, *Applied Statistics for Engineers and Physical Scientists* (2nd Edition ed.). Prentice Hall.

10. Lau, C., 1999, A Comparative Study of Distance Measures Using the Victorian Activity. *21st Conference of Australian Institutes of Transport Research (CAITR).*

11. Ma J. & Goulias K. G., 1997, A dynamic analysis of person and household activity and travel patterns using at a from the first two waves in the Puget Sound Transportation Panel, *Transportation, 24*, 309–331.

12. *Mathwave Data Simulation and Analysis*. (n.d.). Retrieved 2004, from EasyFit – Distribution Fitting Made Easy: http://www.mathwave.com/.

13. McFadden, D., 1973, Conditional Logit Analysis of Qualitative Choice Behavior, in: Zarembka, P. (ed.), Frontiers in Econometrics, Academic Press.

14. Miller E.J., Roorda M. J., 2003, Prototype Model of Household Activity-Travel Scheduling, *Transportation research Record*, 1813, 114-121.

15. Mohammadian, A. and Y. Zhang, 2007. Investigating Transferability of National Household Travel Survey Data. Trasportation Research Record , 1993, 67-79.

16. Stopher P.R., Bullock P., and Greaves S., 2003, Simulating household travel survey data: Application to two urban areas, *82nd Annual Meeting Transportation Research Board, Washington D.C.*

17. Stopher, P.R., Greaves S., and Xu M., 2004, Using Nationwide Household Travel Data for Simulating Metropolitan Area Household Travel Data, *TRB Conference on 2001 National Household Travel Survey, Washington, D.C.*

18. Zhang, Y. and K. Mohammadian, 2008a, Examining Common Distributional Assumptions of Travel Characteristics for Data Simulation. *Transportation Research Board 87th Annual Meeting.* Washington D.C.

19. Zhang, Y., and A. Mohammadian, 2008b, Bayesian Updating of Transferred Household Travel Data, *Transportation Research Record*: Transportation Research Board, 2049, 111-118.