MODELING FRAMEWORK FOR SOCIO-ECONOMIC ANALYSIS OF MANAGED LANES

A Dissertation
Presented to
The Academic Faculty

by

Sara Khoeini

In Partial Fulfillment
of the Requirements for the
Ph.D. Degree in the
School of School of Civil and Environmental Engineering

Georgia Institute of Technology
May 2014

MODELING FRAMEWORK FOR SOCIO-ECONOMIC ANALYSIS OF MANAGED LANES

Approved by:

Dr. Randall Guensler, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael Hunter
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. William Drummond
School of City and Regional Planning
Georgia Institute of Technology

Dr. Michael Meyer
Senior Advisor
Parsons Brinckerhoff

Dr. Michael Rodgers
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: January 6, 2014

To my beloved family

ACKNOWLEDGEMENTS

I would like to express my most gratitude to my advisor, Dr. Randall Guensler, for his continuous guidance, encouragement, support and understanding. His dedication to quality research and teaching, and his professional behavior toward students have inspired me both personally and academically. I would also like to especially thank for the invaluable advice and comments from the other members of my dissertation committee: Dr. Michael Rodgers, Dr. William Drummond, Dr. Michael Meyer, and Dr. Michael Hunter. I would also like to appreciate the entire graduate and undergraduate students who have helped in large-scale data collection and processing. My great appreciation also goes to the transportation group faculties, staff, and graduate students who have created an awesome academic and friendly environment at Georgia Tech.

I am profoundly thankful to my lovely husband, Dr. Hamidreza Marvi. This dissertation would not be possible without his love, support and encouragement. Finally, my deepest gratitude goes towards my mother (Parastoo Nazem), my father (Hossein Khoeini), and my sister (Tara Khoeini). I cannot explain their unconditional and extra-ordinary love and support which is the backbone of my life endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

GP Lane	General Purpose Lane
ML Lane	Managed Lane
HOT Lane	High Occupancy Toll Lane
HOV Lane	High Occupancy Vehicle Lane
HOV-2	Vehicle with Two Occupants
HOV-2+	Vehicle with Two or More Occupants
HOV-3	Vehicle with Three Occupants
HOV-2	Vehicle with Three or More Occupants
HOT3+	HOT Lane Requiring a Minimum of Three Occupants per vehicle
CATI	Computer-assisted Telephone Interviews
CAPI	Computer-assisted Personal Interviews
CASI	Computer-assisted Self Interviews

SUMMARY

Congestion pricing is a fairly recent strategy for mitigating congestion and providing reliable travel options. One form of congestion pricing, High Occupancy Toll (HOT) lanes, utilizes occupancy and toll payment requirements to utilize capacity more efficiently, and to provide consumers with a travel choice that has a shorter and more reliable travel time. Currently, 15 HOT facilities are under operation in the US and more than 16 HOT facilities are under development. Therefore, an understanding of why people choose to use HOT lane is important for policy decisions concerning future HOT lane development (traffic and revenue studies), travel demand modeling, and responding to socioeconomic concerns. Thus far, the travelers' response toward managed lanes is often estimated using stated preference or travel diary surveys, of small percent of the population, which are expensive, time-consuming, and labor-intensive.

How socio-spatial characteristics impact the users travel behavior toward congestion pricing is the main research question of this study. This research is a case study of the conversion of High Occupancy Vehicle (HOV) lane to HOT lane, implemented in 15.5 miles of Atlanta I-85 on Oct, 1 2011. To minimize the cost and maximize the size of the collected data, an innovative and relatively inexpensive modeling framework for socioeconomic analysis of managed lanes has been developed. Instead of surveys, this research is based on the observation of one and a half million license plates, matched to household locations, collected over two-year study period before and after HOV-to-HOT conversion. Purchased marketing data, which include detailed household socioeconomic characteristics, supplement the household corridor usage information derived from license plate observations. The quality of marketing data

is acceptable based on comparative analysis with self-reported survey data. Generalized linear models have been used to link users travel behavior to socioeconomic attributes. Furthermore, GIS raster analysis methods have been utilized to visualize and quantify the impact of the HOV-to-HOT conversion on the corridor commutershed.

At the local level, this study conducted a comprehensive socio-spatial analysis of Atlanta I-85 HOV to HOT conversion to investigate the impact on users' socio-economic attributes and on the commutershed. At the general scale, this study enhances managed lanes' travel demand models with respect to users' characteristics and introduces a comprehensive modeling framework for socioeconomic analysis of managed lanes. The methods developed through this will inform future Traffic and Revenue Studies and help to better predict the socio-spatial characteristics of the target market.

CHAPTER 1

INTRODUCTION

Transportation agencies are faced with growing challenges of congestion and a limited ability to expand freeway capacity due to environmental and societal impacts. Transportation officials are taking advantage of opportunities to address mobility needs and provide travel options through a combination of limited capacity expansion and operational strategies designed to manage travel demand, improve transit service, and support other forms of ridesharing. The managed lane concept is gaining interest around the country as an approach that combines these elements to make the most effective and efficient use of a freeway.

High-occupancy vehicle (HOV) lanes have been in existence since 1969, and the introduction of high-occupancy toll (HOT) lanes in the 1990s has added another alternative for highway management (Fuhs and Obenberger, 2002). Researchers have suggested that in some areas, HOV lanes do not function properly, in that the lanes operate below capacity or the lanes become congested like their general purpose lane counterparts (Guin et al., 2008). HOT lanes promise to make better use of existing HOV lanes, to provide capacity more efficiently than either conventional HOV lanes or general purpose lanes, and to reduce the number of lanes needed on new freeways by managing demand. Other than demand management, HOT lanes provide a choice with reliable and shorter travel time for whoever is willing to pay the toll. In many cases, toll revenues are used for the conversion of a HOV lane to a HOT lane or investment in transit rather than generating funding sources. As of 2013, 18 HOT facilities are under operation in the US, and more than 11 HOT facilities are under construction (Figure 1).

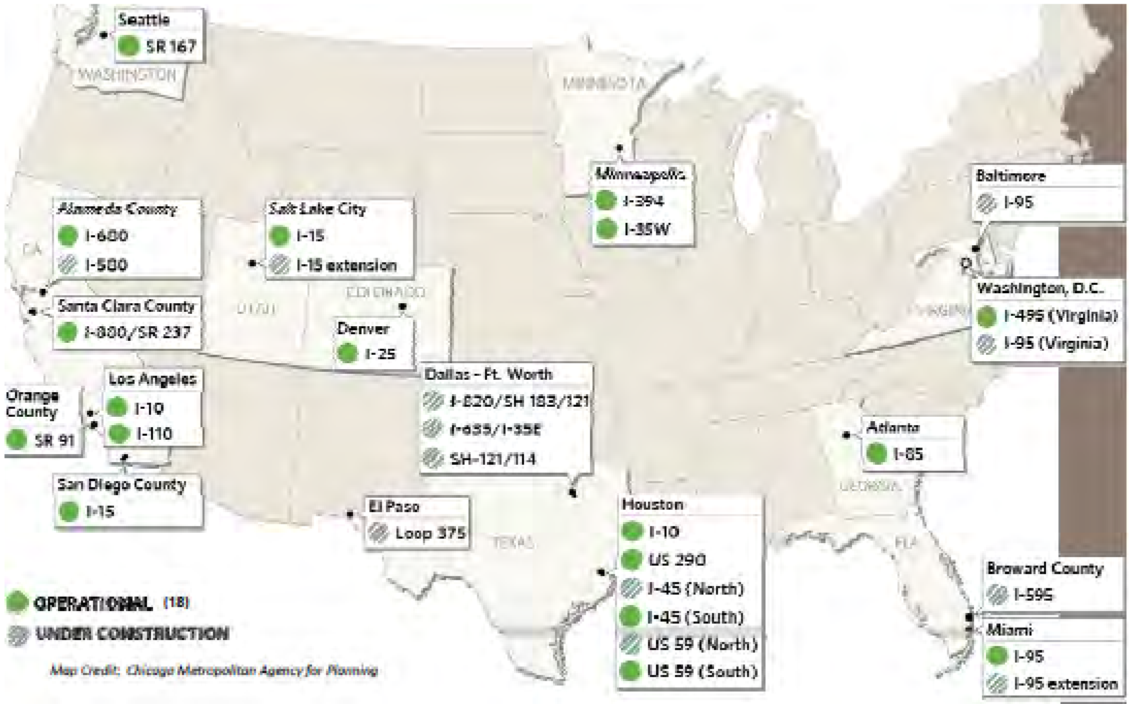


Figure 1 HOV and HOT Facilities in United States (Source: Chicago Metropolitan Agency for Planning)

The conversion of HOV lane to HOT lane has been implemented in 15 miles of Atlanta I-85 (from Chamblee Tucker Road to Old Peachtree Road) on Oct, 1 2011 (Figure 2). I-85 is a heavily commuter-oriented facility, with heavy traffic into (southbound) Atlanta during the morning peak travel period, and out of (northbound) Atlanta in the evening peak. The previous HOV lanes were well traveled during peak periods, that there is little space available to offer to additional vehicles, whether HOV or toll paying. Speeds in the HOV lane drop to 45 mph during the peak hour in the peak direction of travel under previous conditions, or in other words, the HOV lanes were operating roughly at capacity during peak periods in the peak direction.

HOT lane system expansion is under consideration throughout the Atlanta metro region (*HOV Strategic Implementation Plan Atlanta Region*, 2003). Further HOT lanes implementation requires in depth planning and policy analysis and the I-85 experience can help in this regard.

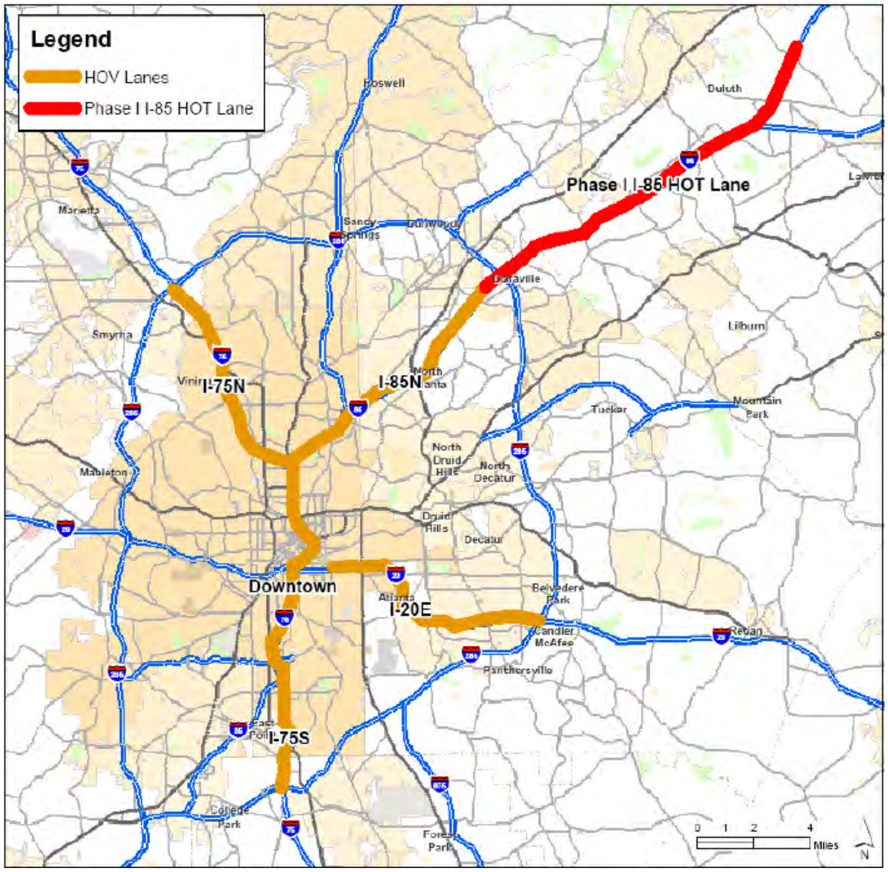


Figure 2 Atlanta Metropolitan Area Managed Lanes Map (source: GDOT Information Center)

Before the Atlanta I-85 HOV to HOT conversion, HOV users had to have two or more passengers per vehicle to be able to use the lane (HOV2+). However, after the conversion to a HOT facility, registered vehicles carrying three or more persons (HOT3+) as well as buses, vanpools, motor cycles, alternative fuel vehicles, and on-call emergency vehicles are allowed to use the lane for free. Because 3-person carpools are more difficult to form than 2-person carpools, the demand for free use of the lane significantly decreases. The resulting capacity on the lane is then sold to 2-person carpools and single occupant vehicles (SOVs) that are willing to pay a toll to use the lane. The dynamic pricing goal that manages the HOT lane demand is to ensure that operational flow is maintained with an average speed of 45 mph or greater, 90% of the time (Priced Managed Lane Guide, 2012).

Before the conversion, people had to choose between general purpose lanes and HOV lane which required two or more passengers per vehicle. After the conversion, solo drivers have a choice of two options: those who choose to pay the toll enjoy a faster trip in the HOT lane, while drivers who choose not to pay the toll may use the congested general purpose lanes for free. Carpoolers on the other hand can use the HOT lane for free if the vehicle meets the minimum occupancy requirement (HOT3+ in I-85). The travel choice before and after the conversion varies by traveler and the same travelers can easily make a different decision on any given trip due to the constraints of that particular trip.

Without any doubt, users' socioeconomic attributes have an important role on their decision making. Unfortunately, the ability to predict and value the users' choice of travel, as a function of socioeconomic attributes, is sparse. Moreover, most of the peer reviewed managed lane travel behavior studies have been looked at travelers' decision making only after the project implementation.

Traffic and revenue studies investigate the feasibility of the project from socio-economic, traffic and operational aspects. However, the majority of traffic and revenue studies are just agency reports prepared only when the final decision for the project implementation has already been made. Furthermore, current analytical models do not capture the full story, regarding HOT lane demand prediction, by only looking at the regional travel demand model. Moreover, in the past, industry experience in the toll demand forecasts upon which these are based have been quite varied, in that demand (and the accompanying revenues) has ranged from overestimated in many cases to occasionally underestimated (NCHRP 364, 2006).

Therefore, not enough research has been conducted yet to link the performance analytical results for previous projects to the future projects traffic and revenue studies especially in terms of socioeconomic impacts and targeting market. This issue becomes much more important considering the funding difficulties, the increasing interest for persuading the private sector for public-private partnership, and therefore the need to financially justify the investment in any particular project.

Thus far, self-reported survey data have been the main sources of data for transportation socioeconomic and travel behavior studies. Depending on the desired level of accuracy, aggregate or disaggregated data have been employed. Aggregate data are released by the Census Bureau publicly and free of charge. American Community Survey (ACS) data (conducted by Census Bureau) are available, to aggregated geographies (Block, Block Group, Tract, etc.) without household level data precision. Household level public data (such as Public Use Micro Sample Data provided by Census Bureau) that are publicly accessible don't have the acceptable spatial accuracy for transportation study purposes.

Other self-reported data sources are disaggregated data in forms of surveys of small percent of the population. Valuable household-level self-reported data, which have been collecting usually in combination with stated preference or revealed preference travel surveys, are expensive, time-consuming, and labor-intensive. To illustrate the costs of household travel surveys, a recent household travel survey conducted by the Atlanta Regional Commission cost $2 million, which was an average of $200 per completed one-day travel survey (ARC, 2011). Meanwhile, the cost and accuracy of household level self-reported data depend upon the depth and comprehensiveness of the surveys that collect the data. Household travel surveys are valued for the amount of behavioral data generated. However, collecting an unbiased and large enough sample of users' characteristics is very expensive and not very feasible for most transportation projects or studies.

Given that self-reported survey data are the basis for previous studies, these surveys appear to lack the ability to model the users' choice toward managed lanes and respond to socioeconomic concerns. The other problem with previous studies is that because they do not have data to study changes in travelers' lane choice (ML vs. GP) in response to pricing during a long enough time period, they lack the power to respond to one main concern, which is assessing any potential disproportionate adverse impacts on low-income populations.

Marketing or credit-report data are alternative sources for household level socioeconomic data (with address-level spatial accuracy) that have not been much the focus of attention in transportation studies. Purchasing a full profile of demographic data costs less than 10 cents per household, also avoiding the time and labor-intensiveness of household travel surveys. Furthermore, marketing data can be updated more frequently (typically by quarter). The potential main reason, for neglecting this informative data source, is the deficiency to provide travel related information such as trips origin, destination, frequency, mode and purpose by household members. To partially overcome these issues, this study collected a very large sample of revealed preference data; more than 1.5 million license plates matched to more than 250,000 households, collected over two years, across different locations along the corridor. A substantial advantage of this study, compared to many counterparts, is the use of revealed preference, rather than stated preference data, from a significantly large sample size.

In summary, an understanding of why people choose to use HOT lane is important for policy decisions concerning future HOT lane investments and developments (traffic and revenue studies) as well as responding to socioeconomic concerns. In spite of previous studies contributions, the emphasis of most of the analysis is not on changes in travelers' choice in response to pricing. Furthermore, the main data sources for previous studies were stated preference or travel diary surveys from a small portion of the population. Research Framework

Figure 3 illustrates the research framework applied in this study. This study started with license plate data collection along five locations in the corridor at morning and afternoon peak hours, over two year study period (one year before and one year after). After license plate data collection, the vehicle registration database has been used to access household locations and vehicle characteristics of the observed license plates. Using the vehicle characteristics, a comparative analysis of vehicle value (as a proxy for income) across the lanes has been conducted. Employing the acquired household locations, spatial analysis techniques evaluated the impact of the conversion on the corridor commutershed. Furthermore, socioeconomic data at household level (marketing data) and at block group level (ACS data) have been retrieved. The quality of marketing data using survey data has been evaluated. Furthermore, the sensitivity of license plate data collection to time, day, and location of data collection has been evaluated using ACS data. Lastly, based on both marketing data and ACS data, socioeconomic analyses have been conducted and managed lane usage models have been estimated.

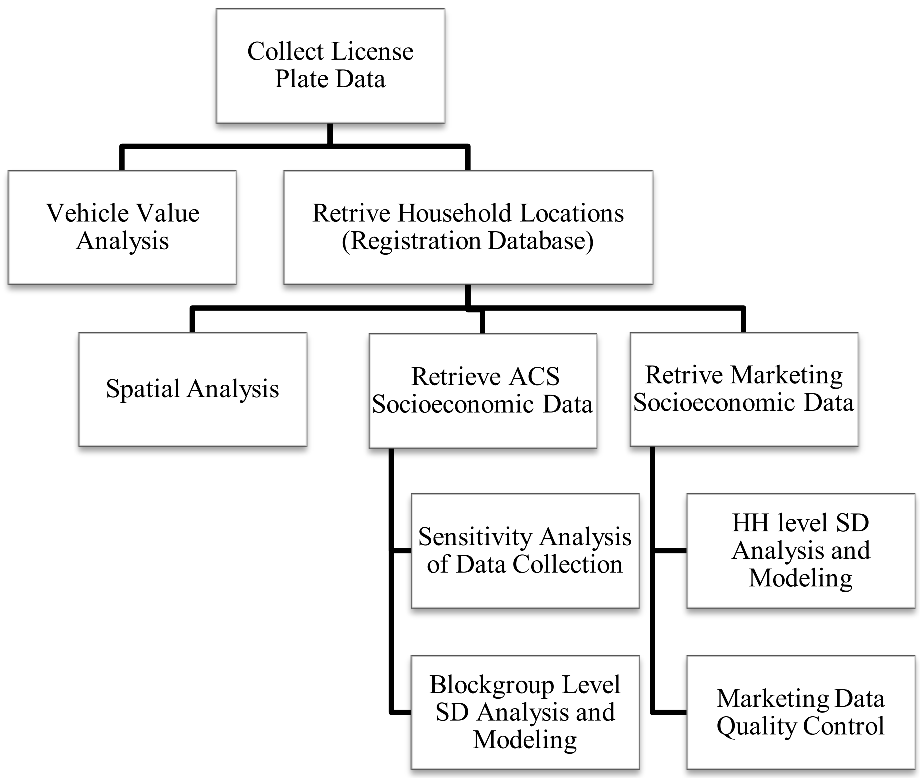


Figure 3 Research Framework

1.1. Research Contribution

Assessing how socioeconomic characteristics impact managed lane usage is the main research question of this dissertation. Specifically, this research will contribute to the state of the art of transportation policy, planning and decision-making in the following four main ways.

1. **Conduct a comprehensive socio-spatial analysis of the Atlanta I-85 HOV to HOT conversion:** This study will analyze the difference between the HOV, HOT and GP

lanes users in terms of socio-economic attributes. The spatial distribution of the commutershed before and after the conversion for HOV, HOT and general purpose lanes will also be analyzed. The main difference between this study and similar previous studies is a very large dataset of users' revealed preference data collected before and after the conversion. This dataset has also created the opportunity to respond to the main environmental justice question, which is assessing any disproportionately adverse impact on low income population; which has not been studied in previous works.

2. **Introduction of a new socioeconomic assessment methodology at the household level with accuracy assessment:** This study introduces credit-report data as a supplement to license plate data for socioeconomic analysis at the household level. This is the first license plate based socioeconomic analysis that uses credit report data. Previous studies either used census data, which are aggregated in some geographic level and lacks variability and individual level accuracy, or surveys which are based on very small percent of the population. The accuracy of this new data source compared to precedent data sources (surveys, aggregated data) will also be examined.

3. **Enhance the managed lane choice modeling with respect to users' characteristics (using I-85 as a case study):** This study models the relationship between the travelers' managed lane use and the socio-economic attributes. Using data collected along two year study period, this study models how travelers' lane choices have changed in response to pricing as a function of socioeconomic characteristics. Different statistical methods will be used, such as cluster analysis, generalized linear

models, multinomial regression, and discriminant analysis to better describe the underling relationship. The large dataset collected from the field, joined to household level credit report data, makes this modeling significantly unique compared to the previous models.

4. **Help future traffic and revenue studies better predict the target market in terms of socio-spatial characteristics of the users:** The findings of the analysis will help future managed lanes toll, revenue and demand analysis to better predict the managed lane usage pattern and so improve the decision making process for evaluating the feasibility of the managed lanes across the region and the nation. Any decision about building a new HOT facility will be able to use the methods developed in this study in the process of establishing and analyzing the socio-spatial aspect of the target market. This research helps ongoing as well as future projects understand potential socioeconomic impacts and plan for preventing or mitigating any negative impacts.

This document is a PhD dissertation prepared in partial fulfillment of the requirements for the PhD degree in the department of Civil and Environmental Engineering, Georgia Institute of Technology. The first chapter explores a comprehensive literature review of all the previous related studies followed by the methodological data collection and processing framework. The next chapter describes data sources applied and collected in this study in addition to quality control assessments. The next two chapters, socioeconomic analysis and spatial analysis, provide the analytical results. Lastly, a modeling framework is summarized and demonstrated for future traffic and revenue studies and a final chapter is presented.

CHAPTER 2

LITERATURE REVIEW

2.1. Congestion Pricing

Traffic congestion is an increasing burden for American cities. Congested highways delay truck transport and commuters, causing economic and social losses to local businesses and residents and making the area less attractive to potential residents, investors, and visitors. Drivers suffer increased stress and the negative health effects. In addition, long delays in car travel leads to greater amounts of pollutants being emitted into the atmosphere, increasing air pollution. Consequently, city and state transportation agencies have a strong interest and great motivation to reduce congestion (Ross et al., 2008).

Congestion pricing (also referred as "Value Pricing") is a fairly recent strategy that has been put forward to mitigate traffic congestion. The concept underlying congestion pricing is one that seeks to manage demand, and thus reduce the number of cars competing for space on the road, by making more explicit the costs of adding an additional driver to the lane. The results are free-flowing travel in the managed lane and a more efficient use of all of the lanes in the managed corridor. Although the number of existing value pricing projects is small, dozens of regions including Atlanta Metropolitan Area are actively considering implementing HOT lanes or variable tolls (*HOV Strategic Implementation Plan Atlanta Region*, 2003).

Current research has found inefficiencies in HOV (High Occupancy Vehicle) lanes and called for eliminating HOV preferential treatment, either by opening up HOV lanes to all traffic or by imposing a toll on all lanes (Dahlgren, 2002). However, supporters of HOT (High Occupancy Toll) lanes argue that the policy promises a number of benefits, including promoting non-SOV modes, especially carpools and transit; generating needed transportation revenues; easing political opposition; and improving air quality (Poole and Balaker, 2005).

The FHWA defines managed lanes as "a limited number of lanes set aside within an expressway cross section where multiple operational strategies are utilized, and actively adjusted as needed, for the purpose of achieving pre-defined performance objectives (Collier and Goodin 2004)". There are numerous benefits of managed lanes to society and users, the main benefits for travelers are travel time savings and more reliable travel times. According to (M. W. Burris et al., 2009), an efficiently operated ML can carry more traffic than a general purpose lane. Thus, MLs provide travel time savings to users and reduce fuel consumption. By reducing the congestion, MLs are also expected to cause less pollution and fewer traffic crashes (Collier and Goodin, 2002).

2.1.1. Congestion Pricing Socioeconomic Assessments

Value pricing is widely recognized to be politically difficult because it adds a price to a public service previously perceived as free (Benjamin and Sakano, 2007). Terms such as "Lexus Lanes" are often used to describe the projects, reflecting a perception that these facilities extend the privilege of a congestion free drive only to those who can afford it (Douma et al., 2005; Smirti and Evans, 2007). Equity concerns frequently center on the question of whether an additional toll would be an unacceptable cost burden for low income communities (Ungemah, 2007).

The basis for examining equity in traditional transportation projects is found in several laws, polices and directives including, but not limited to, the Title VI of the Civil Rights Act of 1964, National Environmental Policy Act of 1969, Executive Order 12898 and the Federal Aid Highway Act of 1970. Specifically, Executive Order 12898 requires that agencies ensure projects are designed to avoid, minimize, or mitigate disproportionately high and adverse human health and environmental effects, including social and economic effects, on minority populations and low-income populations.

Survey results from around the country suggest that people across income groups support the idea of congestion pricing once they are given the choice of selecting a tolled lane with minimum speed guarantees, an alternative route, or a different mode of transportation (Dill and Weinstein, 2007; Ross et al., 2008). A survey of travelers in Dallas and Houston found that the primary reason given for opposition to value pricing was dislike of the tolls, and a view that taxes had already paid for the road was cited frequently (Burris et al., 2007). The view of value pricing as "double taxation" is also shared by many American political leaders (King et al., 2007). However, in public opinion studies support for tolls to fund roadway projects often increases when compared to raising gasoline taxes (Dill and Weinstein, 2007). In the second panel survey of the MnPASS managed lanes on Interstate 394, the most often cited objection to letting single-occupied vehicles (SOVs) use carpool lanes was that only the rich would benefit; the second-most cited objection was that carpool lanes should be free to all (Douma et al., 2005).

Studies on State Route 91 in southern California have shown that at any given time about three-quarters of the vehicles in the toll lanes belong to low- and middle-income individuals with only one-quarter of the vehicles belonging to high-income individuals (Safirova et al., 2003). According to data collected on Express Lanes in California, low-income drivers are as likely to approve of the lanes as drivers with higher incomes. In fact, over half of the commuters (51 percent) with household incomes under $25,000 a year approved of toll lanes.

A 2006 survey on the I-394 MnPASS Express Lanes in Minnesota revealed that MnPASS usage was reported across all income levels, including 55 percent of lower-income respondents (Munnich and Loveland, 2005). The survey also revealed support for the lanes to be high across all income levels, including 64 percent of lower-income respondents. An early look at I-15 FasTrak users found they were more likely to have higher levels of income and education, and to own homes, than non-users (Hultgren and Kawada, 1999). But, surveys have not found differences between higher- and lower income users' attitudes towards congestion pricing (Dill and Weinstein, 2007).

The findings of Atlanta focus group study by Georgia Tech (Ross et al., 2008) before the conversion, shows that there are some concerns about the equity of congestion pricing, and some drivers consider these programs to be taking away a previously "free" service. However, there were no statistically significant differences among socio-economic groups (*e.g.* race, income, education level) in stated willingness-to-pay or willingness-to-use a hypothetical congestion-priced facility. When asked about the possibility of converting existing HOV lanes to managed lanes (any of the three options) there was a concern, especially from current HOV users, that it would constitute taking away something that is currently free. Of those stating they would not use the facility regularly, many indicated they would use it for specific reasons.

2.1.2. Congestion Pricing Travel Behavior Modeling

Thus far, most studies about HOT lanes socio-economic impact have focused on the stated preferences toward HOT lanes (Ross et al., 2008, Burris et al., 2007; Dill and Weinstein, 2007; King et al., 2007; Douma et al., 2005; Munnich and Loveland, 2005; Hultgren and Kawada, 1999; Sullivan, 1998; Supernak and Golob, 2002). For example, the first comprehensive study conducted by Sullivan and his associates (Sullivan, 1998) mainly analyzed the impacts of the HOT lanes on travelers' choice, and assessed public opinions on value pricing. Assessments of SR-91's and I-15 distributional outcomes reveal that low-income drivers do use the HOT lanes, even if less often than wealthier drivers. Furthermore, according to surveys of motorists using both the HOT lanes and the parallel free lanes, low-income travelers expressed approval of the lanes about as often as drivers with higher incomes (Sullivan, 1998; Supernak and J. Golob, 2002).

Although stated preference surveys have been the main basis for the pricing socioeconomic analysis, they have mainly conducted before the project implementation and in some cases did not match with the revealed preferences (Hultgren and Kawada, 1999; Munnich and Loveland, 2005). Moreover, the sample sizes are very small and based on the basic method for collecting stated preference data may be biased. The other problem with these studies is the lack of statistical analysis component in modeling travelers' decision making in response to the pricing.

Recently, a few studies have touched upon the factors that are associated with HOT lane usage, using statistical methods. However, almost all of the studies conducted statistical modeling based upon the stated preference/travel diary surveys from a small sample of the population. For example, Li, et al., (2007) examined the determinants of

HOT lane use with the first comprehensive survey data (sample size = 759) on the State Route 91 Express Lanes in California using multivariate logistic regression. The results showed that controlling for other variables, household income, vehicle occupancy, commute trip, and drivers' age were important predictors of high-occupancy-toll lane use, but gender, trip length, trip frequency, and other household characteristics make no significant differences in high-occupancy-toll lane use (Li, 2007).

Another study by Burris (2012) used mixed logit model to examine the impact of several personality traits on survey respondents' travel choices. Burris found that several variables particularly travel time, toll, gender, and income, were better predictors of ML usage than the psychological variables such as conscientiousness, locus of control, risk attitudes, personal need for structure, driving risk perceptions and driving styles (Burris et al., 2012).

Another study showed that the most important predictor on managed lane usage was income, with higher income travelers more likely to use the managed lane; however, travelers from all income levels will use the managed lanes (Devarasetty et al., 2012, Devarasetty, 2012). The same study showed that females tended to use the managed lanes more often than males. Also, users who were late for an appointment, have an important appointment, or are worried that they will be late if using the general purpose lanes will have the highest willingness to pay for the managed lanes (Patil et al., 2011).

There are also some shortcomings with these recent lane usage studies. Similar to other socioeconomic impact assessment studies, the studies discussed previously are all based on very small samples, model goodness of fit is low, and study results are not consistent across different studies. For example, while income was consistently significant in predicting managed lane usage across all of the studies, the magnitude of the predicted impacts differed considerably. Moreover, age and gender were not significant across all of the studies (Burris et al., 2012; Devarasetty et al., 2012; Li, 2007). Other critical socioeconomic variables, such as ethnicity, have never been identified as significant in previous studies. More importantly, because these studies did not have data to study changes in users' choice in response to pricing during a long enough time period, they lack the power to respond to the main environmental justice question which is looking at the disproportionately adverse impact across demographic groups.

The main data source for the development and calibration of the Atlanta travel demand model used in the traffic and revenue study for Atlanta was a household travel survey of eight thousand households, conducted for the Atlanta Regional Commission (ARC) from April 2001 through April 2002 (Jacobs, 2009). However, the HOT lane usage patterns are likely to be significantly different from general corridor usage patterns. Moreover, the socioeconomic analysis was conducted at the county level by simply assessing county-level socio-economic characteristics and their trends over time, without any link back to previous projects. Such limited methods do not provide a justified argument for the potential market share of a managed lane.

Use of the standard travel demand modeling approach to forecast demand for the HOT lane under operating conditions that included pricing, and market sector response to pricing, was apparently inadequate. This is evident by the observation that, to date, demand still exceeds capacity under the maximum toll price on the Atlanta I-85 HOT corridor. Unfortunately, not enough research has been conducted yet to link the previous projects performance analytical results to the future projects traffic and revenue studies especially in terms of socioeconomic impacts and targeting market.

2.2. Travel Survey Methods

2.2.1. Function of Socioeconomic Data in Transportation

Socioeconomic data sources are needed for various types of transportation-related socioeconomic assessments. Modeling of socioeconomic impacts is unlikely to be reliable unless derived from reasonably large samples of representative data.

2.2.1.1. <u>Equity Assessment</u>

The urban area is dealing with challenge of sustainability which achieved by respecting the equal distribution of quality of life measures across various demographic groups. *Social equity* (also called *fairness or social justice*) refers to the equitable distribution of impacts (benefits, disadvantages and costs) (Litman, 2002). This is an important planning goal, and a requirement for sustainable development which considers social as well as economic and environmental objectives (Litman and Burwell, 2006). Sustainable development requires more integrated planning that considers a wider range of impacts and options, identifies and implements *win-win* solutions, that is, policies and programs that help achieve economic, social and environmental objectives. The most important variables in equity studies are age and life cycle stage, household type, income, race and ethnic group, and disability related measures (Litman and Brenman, 2012).

2.2.1.2. <u>Travel Demand Modeling</u>

To make informed transportation infrastructure planning decisions, planners and engineers need to be able to forecast the response of transportation demand to changes in the attributes of the transportations system and changes of the attributes of the people using the transportation system (Bhat and Koppelman, 1999). Travel demand models are used to predict travel characteristics and usage of transport services under alternative socioeconomic scenarios, and for alternative transport service and land-use configuration. The decisive variables in travel demand modeling other than travel behavior variables includes but are not limited to household size, age, vehicle availability, income, license holding, gender, employment, education, dwelling type and dwelling ownership (Kitamura et al., 1997).

2.2.2. Review on Travel Data Collection

In the early days of transportation planning, survey data were collected using face-to-face home interviews, with sample sizes that were often as large as 1– 6% of the population (Meyer and Miller, 2001). The first major technology utilization in travel survey was the use of telephone and computers to collect data called computer-assisted telephone interviews (CATI), computer-assisted personal interviews (CAPI), and computer assisted self-interviews (CASI). The household socioeconomic data in addition to trip making attributes were collected simultaneously in these aforementioned methodologies (Meyer and Miller, 2001).

Over the past decade, the survey sample sizes have dropped considerably, and are more often now in the range of 2500–10,000 households, representing less than 1% of households in the region. Furthermore, two-day travel surveys have been reduced to one-day travel surveys. Moreover, surveys have become both too expensive and sensitive considering privacy aspects (Stopher and Metcalf, 1996). For example, Atlanta Household Travel survey conducted in 2011 cost two million dollars for collection of about 10,278 households socioeconomic and trip data, which represents less than 0.5% of the metro area population (ARC, 2011). A household travel survey by the Durham-Chapel Hill-Carrboro Metropolitan Planning Organization required $208 per completed survey (Rousseau, 2011).

Another trend in household travel surveys is increasing non-response rates (Wilson, 2004). Furthermore, it is known that many of the households that are non-respondents travel more than the average, or are larger households (NCHRP, 2006), which potentially creates bias in the collected data. The fact that recent technology advances such as smart phones or high speed Internet may not be equally distributed among the population (income groups, age groups, etc.) could also introduce sample bias accordingly. Furthermore, the process of conducting surveys is very labor intensive.

The next major enhancement has been the use of passive location data by applying Global Positioning System (GPS) data loggers either in vehicle or hand-held devices. The latest advances technologies are GPS-enables smart phones and RFID tag reads (Doherty, 2009). Electronic toll collectors identify the user by reading the user's RFID toll tag.

Active/interactive technologies such as computer user interfaces and cell-phone apps have initiated the collection of socioeconomic attributes as well as detailed trip characteristics. A joint project by researchers at IBM and MIT (Lorenzo and Reades, 2012) concluded that fine-grained, extensive data from mobile phone networks "is providing us with a more comprehensive view of activity and mobility at the urban scale than travel diaries can possibly do on their own. It also enables us to shed light on hitherto invisible intra-personal variation in travel activity."

Cellular technology forgoes sensors and detects spatial data from mobile devices, such as cell phones and tablets, as each one contacts its cellular network. When users make or receive calls, send or receive texts, browse the Internet, or send or receive email, their approximate spatial locations could be saved and integrated to build the trip making patterns. Compared to data gathered from household travel surveys, cellular technology provided researchers with information about individual mobility with a lower collection cost, larger sample size, higher update frequency, and broader spatial and temporal coverage (Wang et al., 2012).

The cost of collecting cell phone data is relatively low. A recent study by the town of Sierra Vista, Arizona, measured travel across 80 districts, covering 16,000 square miles for 12 weekdays, and collected cell-phone data on more than six million trips for $10,000 (AirSage, 2013).

Lastly, by using high-speed cameras or ALPRs, license plate surveys attempt to recognize license plate numbers and then compare those numbers to state motor vehicle registration databases to identify the households (Colberg, 2013). Accordingly, rate of observing license plates in each lane, route, and time could provide valuable information about users travel behavior (Khoeini, et al., 2012).

Whereas all the recent information technology advances and technological innovations produce accurate and large samples of trip information, they lack the valuable socioeconomic piece of information. Marketing data have been introduced in airport trip generation studies for the first time (Kressner and Garrow, 2012) as a potential household level socioeconomic data source.

Marketing companies collect household/individual-level data using credit reports and any self-report data that they can discover and assign to household addresses with associated attributes. They also utilize imputation models for missing variables or households. However, the imputation model and their data sources remain proprietary and confidential. These marketing companies compete on the availability, accuracy and cost of their data. There are diverse sources available for purchase and the price depends on the size and attributes of the requested data. The average cost of purchasing marketing data is less than 10 cents per household, which is significantly lower than the household travel survey cost including the cost of supplemented travel data (license plate data, cell phone data, etc.).

CHAPTER 3

DATA COLLECTION

3.1. License Plate Data collection

A performance evaluation of the I-85 HOT lane was conducted, from 2010 to 2012, by a team from Georgia Tech -School of Civil and Environmental Engineering- to assess the impacts of the congestion pricing on facility throughput and commuters' demographic profiles. Quarterly vehicle occupancy (persons/vehicle) and license-plate data were collected for one year before and one year after HOT lane implementation. The collected data were coupled with traffic volume data and used to evaluate the effectiveness of the HOT lane and to aid in the assessment as to whether the HOT lanes should be expanded to other corridors in the Atlanta area. Data were collected at the following locations:

- Chamblee Tucker Road (CTR) located at I-85 Exit 94
- Jimmy Carter Boulevard (JCB) located at I-85 Exit 99
- Beaver Ruin Road (BRR) located at I-85 Exit 102
- Pleasant Hill Road (PHR) located at I-85 Exit 104
- Old Peachtree Road (OPR) located at I-85 Exit 109

The locations of the sites can be seen in Figure 4. A total of eight quarterly deployments were conducted in fall, winter, spring and summer (starting in fall 2010 and ending in summer 2012). During each quarter, a team visited each of the five selected sites for three week days for the starting and the ending sites and two week days for the three middle sites to collect license plate and occupancy data in both the AM and PM peak periods. Each peak session collected data for two hours: 7am-9am for the AM-peak and 4:30pm-6:30pm for the PM-peak. Because traffic around the Atlanta area generally enters the city in the morning and exits the city in the afternoon, the AM-peak sessions observed the southbound traffic while the PM-peak sessions observed the northbound traffic.

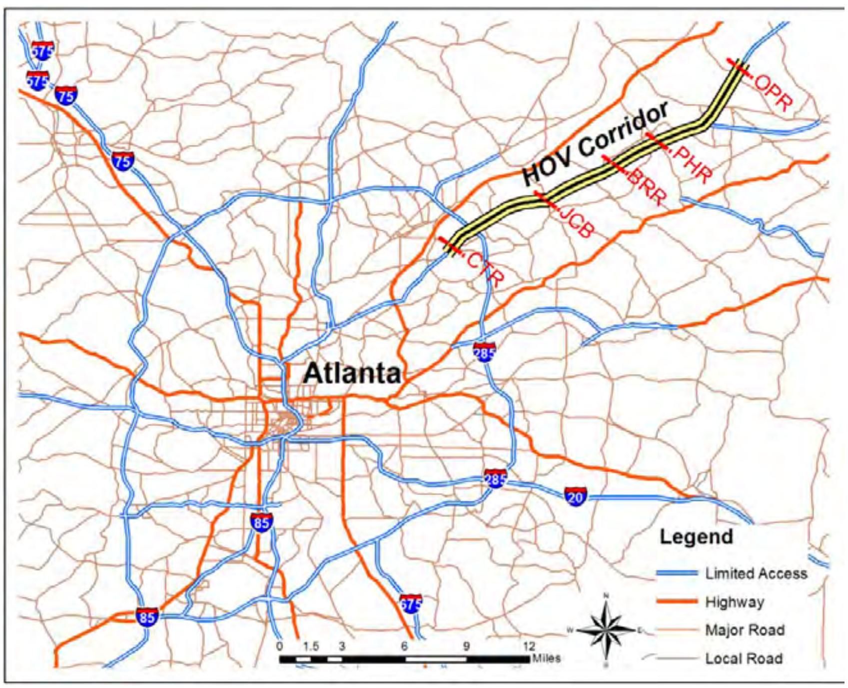


Figure 4 Atlanta I-85 HOV-to-HOT Corridor Data Collection Sites

3.2. Linkage to Household Addresses

After data collection, the recorded video was manually processed using proprietary video processing software developed at Georgia Tech. The processed output file includes license plate characters, date, time, and site of data collection. The method of converting video files to license plate numbers is based on visual capture by undergraduate students. If the license plate is unreadable, the processor records the vehicle as “missed” to allow for a reliable vehicle count. There are several factors that can create a missed record while the video is being processed. Low light levels, blurred video, tailgating, towing, and lane changes were the most common reasons for a missed license plate. During the fall and winter quarters, sunset and sunrise during the data collection sessions were causing low light levels, which affects the quality of the HD recording and thus the visibility of the license plate.

With this methodology, vehicle license plate identification rates range from a low of 50% under poor lighting conditions to a high of 95% under ideal conditions (D’Ambrosio, 2011). During data collection periods with reasonable lighting, typical capture rates were on the order of 70% to 80%. The OPR site had more missed license plates compared to the other sites. Most of these missed records occurred during AM data collection which was immediately after a daylight saving time change and caused very poor lighting condition during the morning. Because peak hour traffic was not based on lighting conditions, data collection times didn’t change. The average capture rate of the data used in this study was 75%, an average value for all five sites. Details on the license plate data processing procedures can be found here (D’Ambrosio, 2011). Because the study only covers GA registered vehicle, out of state vehicles were excluded in the first step of data processing, eliminating approximately 5% of the collected data.

The next step was matching the license plates with Department of Motor Vehicles Database. Because the license plate number collected were not 100% accurate due to environmental condition or human errors, only the matched license plates which were matched to addresses in the Georgia registration database were considered in the analysis. To address privacy concerns, the matching process to the registration database was done on a remote machine, by a third party, using a key ID. Therefore, database tables never had both the household addresses and the license plates at the same time.

It is important to note that not all the commuters observed actually live in the place that they have registered their vehicles (Nelson et al., 2008). Therefore, the registration address may be different from the actual residential address. For example, students and young professionals may register their vehicle at a parent's address to reduce their insurance rates. Couples may live together in the corridor, while maintaining separate addresses. Governmental and commercial license plates account for approximately 10% of the vehicles in our study. Users of these vehicles may use these vehicles for their daily commute trips. Leased vehicles are usually registered by the car owner household address instead of leasing company address, but not always.

Based on a study (Granell, 2002) 33% of the total vehicles were not registered in the same place that they start their daily trip. However, based on GIS spatial tools, 87% of the registered vehicle addresses did fall within the Atlanta metro area. Gwinnett County alone represented more than 66% of all the license plates. The next highly most-involved counties are Fulton and DeKalb County; the other two large counties adjacent to the corridor. After joining to the registration database and geocoding the addresses, 53% of the observed license plates in the field could be matched to a valid location in Atlanta metro area, which assumed as the basis for this study.

3.3. License Plate Data Summary

During seven seasons of data collections (170 two-hour sessions) about one and half million valid license plates (matched to vehicle registration database) were collected which creates a unique database of about 280,000 households. The average frequency of household observation for the households observed is close to five over 170 sessions of data collection.

Before the conversion, license plates belonging to 196,504 unique households were observed. After the conversion, license plates belonging to 166,712 unique households were observed. The number of households observed after the conversion decreased due to a reduction in data collection (explained in section 5.1.4). All of the observed license plates created a database of 278,517 households. On average, the team observed license plates from each unique household approximately five times over the two years study period. The database was used to identify frequent corridor commuters based on the frequency of observation.

Figure 5 shows how frequent the observed households are distributed during one year before and one year after the study. The top-left figure shows the households observed before the conversion across all the lanes including the HOV lane (n = 196,504). The average observation frequency per household, before the conversion, is three. Similarly, the top-right figure shows the households observed after the conversion across all the lanes including the HOT lane (n=166,712). The average observation frequency per household, after the conversion, is close to two. Finally the bottom figure shows all the households observed on the corridor both before and after the conversion (n=278,517). The average observation frequency is five.

Figure 6 shows how frequent the observed households are distributed only along the HOV/HOT lane during one year before and one year after the study. While, 20% of the total observed households have at least one observation in the managed lane, only 2% of the observed households have both before and after observation in the managed lane.

14% of the households have been observed in the HOV lane and 4% of the households have been observed in the HOT lane.

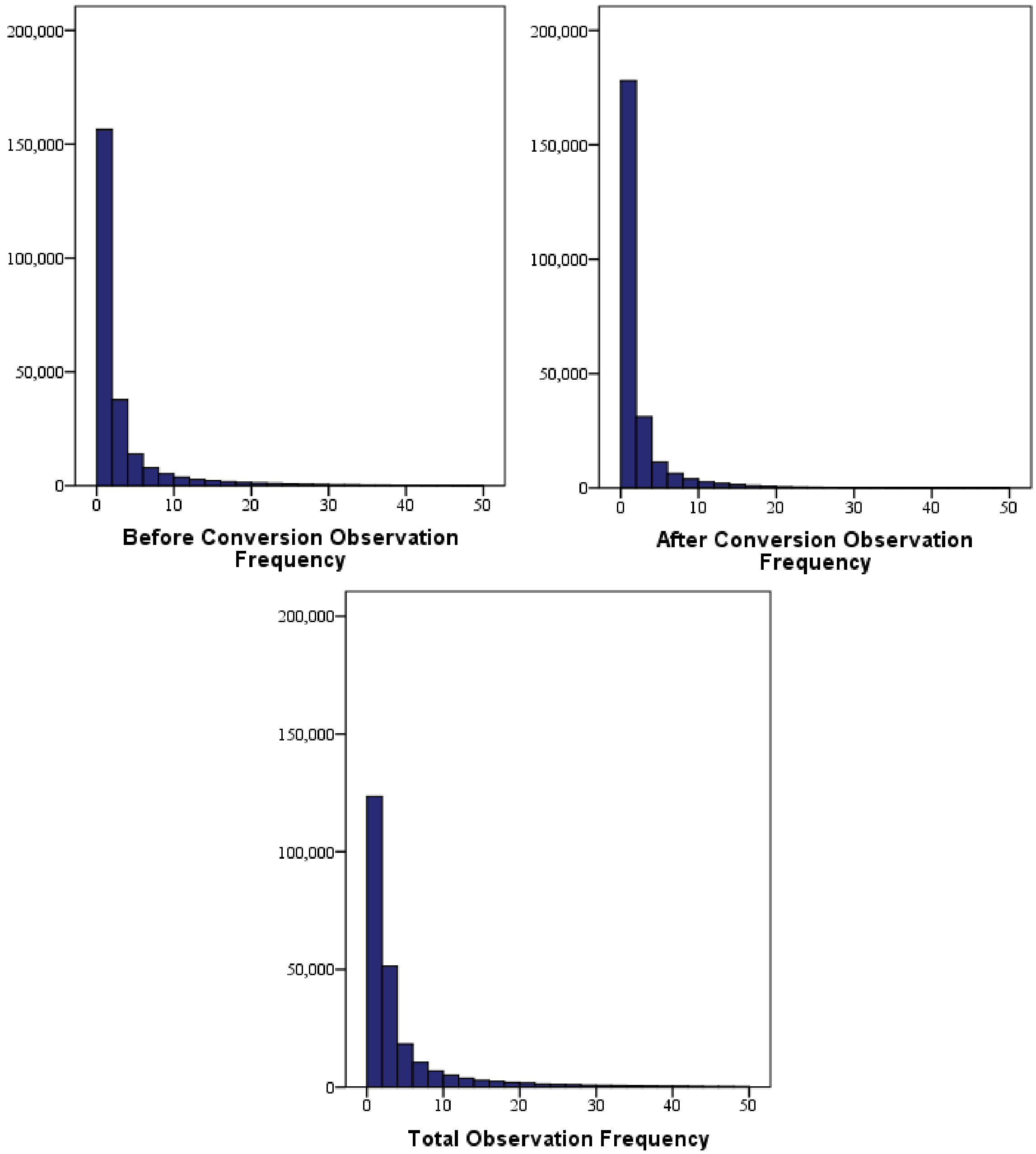


Figure 5 Histograms of Households Observation Frequency

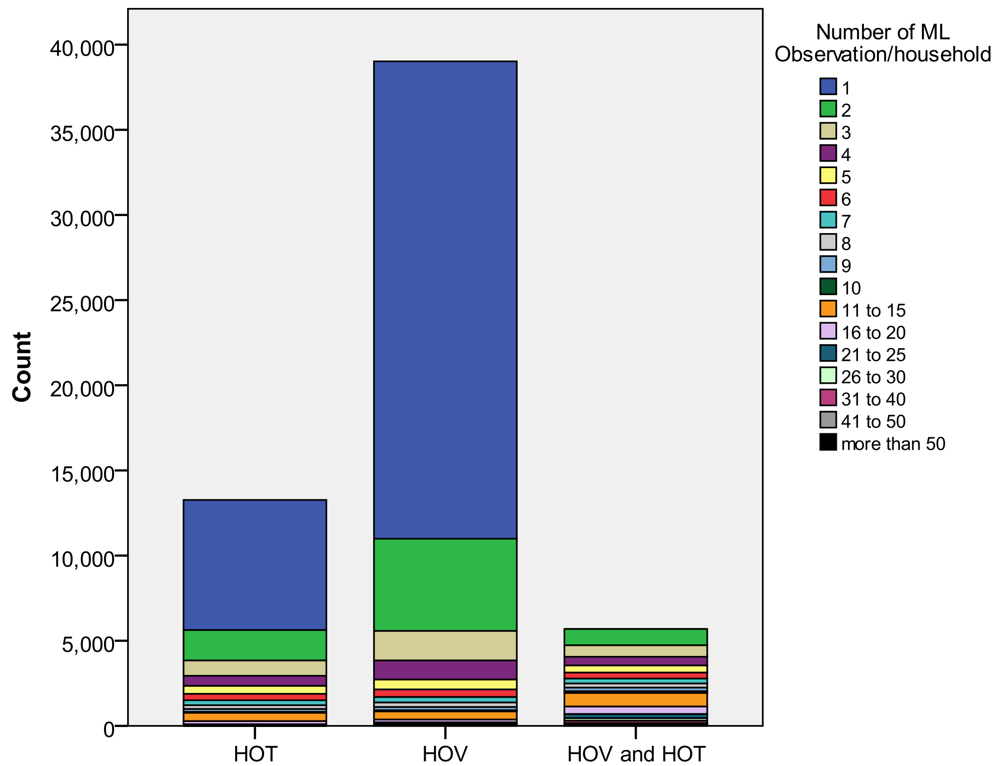


Figure 6 Observed Households along I-85 HOV/HOT Lane Color-coded by Frequency of Observation

3.4. Managed Lane Usage Frequency

Figure 7 illustrates what percent of managed lane trips are conducted by what percent of the households. Whereas 22% of corridor users were responsible for all HOV trips, only 12% of corridor users are responsible for all HOT Trips.

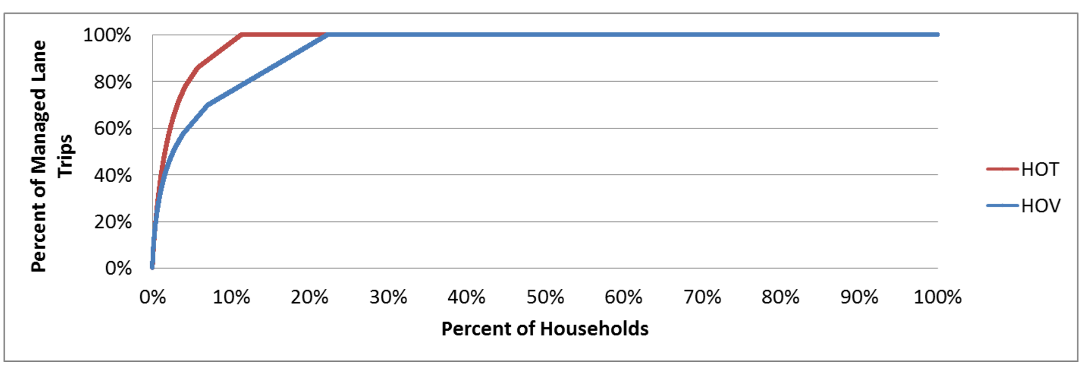


Figure 7 HOV and HOT Frequency of Use

To further investigate the frequency of managed lane usage across the users, households with more than 5 frequency of observation before the conversion for HOV lane usage distribution and households with more than 5 frequency of observation after the conversion for HOT lane usage distribution have been selected. Figure 8 illustrates the ML-usage distribution charts. Since there are so many households with zero percent managed lane usage, these figures removed the non-ML users to better visualize the distributions.

Comparing HOV to HOT usage, HOT usage distribution looks more like a uniform distribution with average equal to 0.50 excluding the non-ML users. However, HOV usage distribution looks more like an exponential or gamma distribution with average equal to 0.28 again, excluding the non ML users. The non-ML users for HOV correspond to 63% of the HHs while this number increased to 70% for non-HOT users; this was actually expected because of the requirement to buy a transponder to be able to use the lane for all the motorists.

Specifically, 20% of the households were using HOV lane 0% to 20% of the times, while the same number equals to 7% for the HOT lane. On the other hand, 18% of households are using HOT lane more than 80% of the time while only 6% of the households were using HOV more than 80% of the time. Therefore, distributions of managed lane use, comparing HOV lane to HOT lane, have changed significantly.

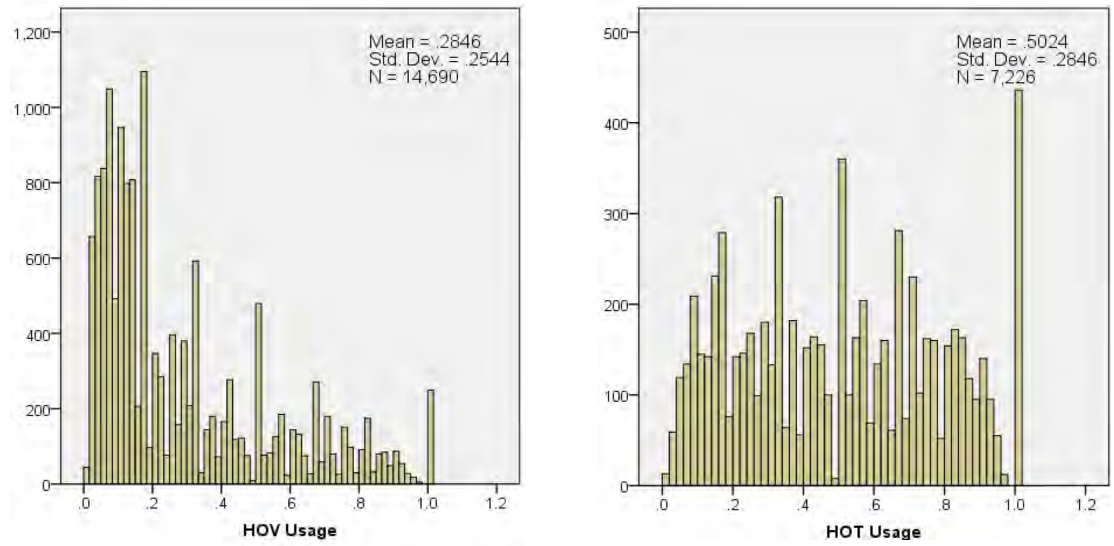


Figure 8 Managed Lane Usage Distribution Charts (Left: HOV, Right: HOT)

3.5. Link license plate records to socioeconomic data

Georgia license plates (over a million and a half) were matched to socioeconomic data using a step-wise process, using a key identifier rather than physical address data for privacy considerations. The key identifier was carefully designed to include information about the plate observation (site, session, period, etc.). For each license plate, the household location was identified in the registration database, linked via the key identifier to the field record, socioeconomic characteristics for the household were appended to the record using census and private marketing data sources, the household location data were removed, and data were returned for analysis.

CHAPTER 4

DATA SOURCES

4.1. Socioeconomic Data Sources

This chapter explains the various data sources were used in this study including American Community Survey, Vehicle Registration Database, Household Travel Survey, and Marketing Data. Moreover, the quality of marketing data was assessed using household travel survey data and is presented in this chapter.

4.1.1. American Community Survey

The US Census Bureau provides publicly available household socioeconomic data aggregated by geographic boundaries (block groups, tracts, counties, etc.). Until 2000, household socioeconomic data were collected through decennial census long from surveys from about one in every six households. However, starting in 2005, the American Community Survey (ACS) has been collecting household socioeconomic data annually. ACS is a part of the U.S. Census Bureau's Decennial Census Program and is designed to provide more current demographic, social, economic, and housing estimates during the decade between Census data collection.

Each year, the survey randomly samples around 3.5 million addresses (1% of total US addresses) and produces statistics that cover 1-year, 3-year, and 5-year periods for geographic areas in the United States and Puerto Rico. The 5-year estimates are available for distinct geographies including the nation, all 50 states, DC, Puerto Rico, counties, places, Census tracts, and Census block groups. The ACS Summary File data cover demographic, social, economic, and housing variables. The ACS 5-year estimates contain additional summary levels, such as census tracts and block groups that are not published in the ACS 1-year and 3-year estimates (*5-Year Summary File Technical Documentation*, 2012).

Because of the small sample size, blockgroup level (smallest available level) household data are provided from ACS using aggregated 5 years of data (5% household coverage). Therefore, in this study ACS 5-year summary file 2005 to 2009 (available at the time of analysis) have been used.

4.1.2. Vehicle Registration Database

State vehicle registration database contains detailed information about the vehicles registered at each household such as household location, number of registered vehicles and their make, model, and model year. However, considering privacy aspects special procedures are employed to detach license plate observations from household addresses used in the analysis. These procedures were implemented on a third party machine using a step-wise process to prevent restructuring the registration database. Therefore, actual license plate characters, household locations, and vehicle characteristics have never been presented in a same dataset. Considering the presence of business-owner license plates, all the addresses with eight or more vehicles registered at their address have been removed from the analysis.

4.1.3. Household Travel Survey Data

While stated preference or travel diary surveys may provide accurate self-reported data from the managed lane travelers, their sample size typically only covers small percentage of the usual population. However, license plate data collected during more than 300 hours of data collection creates a comprehensive dataset of the users and any analysis based on this large dataset (assumed to include accurate socio-economic data in household level) is statistically more reliable. Survey data were used to evaluate the accuracy of the household level (from third party sources) data. The surveys which will be used in this study are explained in the following sections. The socio-economic data have also been purchased for the households of these two surveys respondents for quality control purposes.

4.1.3.1. ARC Household Travel Survey

The Atlanta Regional Commission (ARC) conducted a comprehensive study of the demographic and travel behavior characteristics of residents within the 20-county study area. The purpose of the 2011 Regional Travel Survey was to improve the ARC travel demand forecasts, in both its aggregate four-step trip-based model and its disaggregate activity-based model. The goal was to obtain demographic and trip data from a minimum of 10,000 households, including a subsample of 1,000 households that would also provide global positioning system (GPS) data. The final data set contains information for 10,278 households, of which 1,061 households also provided GPS data.

4.1.3.2. Volpe Household Travel Survey

To support the Federal Highway Administration, the Volpe Center administered a household travel survey as part of the evaluation of the Atlanta Congestion Reduction Demonstration (CRD) Program (Greene et al., 2012). Specifically, the purpose of the survey was to better understand the impacts of pricing on travel behavior on the I-85 corridor. A panel approach was utilized, in which the same households were surveyed both “before” and “after” the deployment of pricing.

The survey consisted of a household-level demographic questionnaire, a 2-day travel diary, and follow-up questions on general travel patterns and travel-related attitudes. Nearly all respondents completed their surveys using a web-based interface (a toll-free telephone number was also available). Wave 1 (“before”) data collection was conducted in April-May 2011 and the Wave 2 (“after”) survey was conducted in April-May 2012. The HOT lane opened on October 1, 2011, about halfway between the survey waves.

Overall, 1,655 households (3,126 individuals) completed both waves of the survey. The population of interest was defined as peak period users of the I-85 corridor and all adult members of their households. Households invited into the survey process were identified via license-plate capture (photography) during peak hours on sections of I-85 and Buford Highway. The license plate numbers were then matched to vehicle registration data and a survey invitation was mailed to the household. Because the survey’s sampling plan was stratified by route and mode, the data were weighted at the household-level and person-level using data for peak period travel volumes in the I-85 corridor by route and mode. Approximately, 95% of respondents used online survey tool, but the telephone option was available and used by about 5%.

The socioeconomic attributes of the respondents such as income, gender, age, race, education, household composition, employment status, technology ownership and vehicle ownership have been collected. Additionally, two-day travel profile including travel mode, trip purpose, travel time, trip tours, departure and arrival times, occupancy, costs, route and any detail information about the trips they took have also been collected.

Moreover, some questions about their attitude toward the HOV and HOT lane have been asked from the participants.

4.1.4. Marketing Data

After thorough research on publicly available demographic data, the research group came to the conclusion that more accurate data sources are required to analyze the demographic aspect of I-85 HOV to HOT conversion. Following paragraphs summarizes some underlying reasons for this decision.

After Census 2000, Census Bureau stopped deploying the long form Census which collected detailed household characteristics from 1 out of 6 households is the United States. Beginning in 2005, American Community Survey (ACS) started collecting the same detailed household characteristics from only one percent of population every year. Because of the small sample size, the lowest aggregation geography level for which ACS data is releasing is Block Group aggregating 5 years of data (2006 to 2010 available now). While the data for this study were mainly observed in 2011 and 2012, the available ACS data were somewhat outdated, especially considering the economic downturns, in previous years.

Furthermore, the variability of household characteristics in a block group is high, making one mean/median value is non-representative for all the households (600 households per block group in average) in one blockgroup. This fact becomes obvious by comparing your household characteristics to your neighbors. In summary, one aggregated value over a 5 year period for 600 households is simply not enough for managed lanes socio-economic analysis.

After investigating all available household level data sources, we found that credit report companies provide all detailed information of the households appended to their household address or householders' names. Credit bureau data cover all detail aspects of the household information such as household size, composition, age, gender, income, education, occupation, dwelling type, ethnicity, life style, home ownership, home characteristics and automobile ownership. Marketing agencies have different sources for their information including self-reported data and they include the source of information in their dataset (whether it is from self-reported data or other sources) for any piece of information.

To better illustrate the difference between aggregated and disaggregated data sources, Figure 9 shows the income distributions for about 2,500 households observations along HOT lane using two data sources. For aggregated data source ACS 2005-2009 5-year aggregated data in blockgroup level has been used. For disaggregated data source credit report data purchased on 2010 has been used. Although the income classification for both of the histograms are the same (the classification boundaries are illustrated in the figure), the distributions are significantly different, as well as the mean value. Because of the aggregation of ACS data, most of the cases are concentrated in the middle around class six or seven. However, in the credit-report data histogram, the cases are deviated more from the center. Although household level data illustrates higher accuracy, the analysis will both include aggregated and disaggregated data to understand the impact of different data sources.

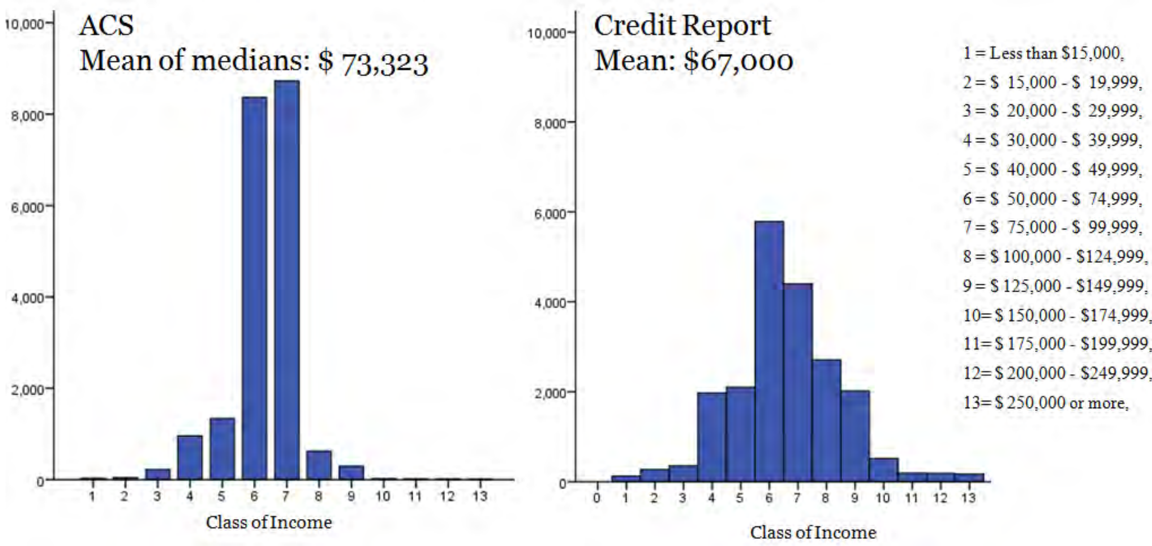


Figure 9 Income Distribution Using Aggregated vs. Disaggregated Data (Left: Block groups' Median Household Income, Right: Household Level Income)

4.2. Marketing Data Quality Assessment

Household travel surveys have been critically important in transportation. MPOs rely on carefully planned and implemented household travel surveys to build regional travel demand models. Household travel surveys involve sending surveys (via mail, phone, or e-mail) to carefully selected households. Participants are typically asked to log the day, time, duration, mode and purpose of every trip for a few days in addition to socioeconomic and demographic information. Household travel survey data are valuable because they provide both trip information (purpose, mode, time, length, etc.) as well as the socioeconomic information of the households in one dataset.

As valuable as household travel survey data are, the amount of data is limited. The cost and significance of self-reported survey data are dependent upon the size and comprehensiveness of the samples. After all, collecting an unbiased and large enough sample of users' characteristics across the wide variety of demographic variables of concern in transportation planning is very expensive and not feasible for any project or case study. To illustrate the costs of household travel surveys, a recent household travel survey conducted by the Atlanta Regional Commission, that collected data from 10,000 households, cost $2 million (ARC, 2011).

Household travel surveys usually collect travel information for only one or two days from a sample of the population. Furthermore, the accuracy and inclusiveness self-reported responses could not be evaluated. Household travel surveys are mainly available for metropolitan areas, with less concentration on regional or rural trips. Lastly, the fact that household travel surveys collect data for a very short time period, eliminate the ability to look at longitudinal changes across the time especially for before and after studies such as congestion pricing evaluations (Schönfelder et al., 2006).

Recent informational and technological advances have created a wide range of potential data sources on travel demand and traffic information, such as spatial trip data through GPS loggers and cell phone traces, RFID tag readers, and license plate data collected via automatic license plate recorders (ALPRs). While these advanced technologies collect spatial data, they are lacking the socioeconomic piece of information which is critical for transportation studies.

Marketing or credit-report data are a potential alternative source for household level socioeconomic data (with address-level spatial accuracy), but have not been the focus of much attention in transportation studies. Marketing companies collect household/individual-level data, using credit reports or any relevant data source, and assemble a database of household addresses with associated attributes. These companies compete on the availability, accuracy and cost of their data. Numerous marketing database resources are available for sale, and the price depends on the number of records purchased and attributes of the data. However, the cost of purchasing general demographic data is typically less than 10 cents per household.

The main purpose of marketing data is to assist business owners, companies and industries to more precisely and efficiently acquire their target market and therefore minimize their costs and maximize their revenues. Thus, access to accurate and informative household information has a great value in today's business environment and therefore there is a substantial willingness to pay for such information. This fact has motivated marketing agencies to enhance the quality, timeliness, extent and coverage of their data to be able to compete with other agencies. Hence, wherever certain variables are not available for some households, marketing agencies develop and apply imputation models to fill out the empty records and improve their coverage. Marketing agencies also provide trigger variables which illustrate if a variable has generated from these imputation models for any household.

The research team compared data for several marketing agencies prior to making a data purchase decision. Considering the non-advertising role of academic publications, no information about the marketing agency and related claims regarding the quality and coverage of their data will be discussed here. In fact, the purpose of this section is to evaluate the marketing data potentials in transportation studies, regardless of the entire commercial claims about the data quality and coverage. All the available socioeconomic variables provided by the marketing agency have been purchased for 10,000 households. To address privacy issues, the 10,000 Atlanta household travel survey addresses were combined with more than 200,000 addresses across the Atlanta metro area for the data purchase. Albeit, different variables have different coverage ratio and not all the variables are available for all the households.

The marketing data do not provide travel behavior information and cannot be substituted for travel behavior data collection via travel diary surveys in some form. Hence, purchased demographic data are rarely employed in regional travel demand studies. However, capture of license plate data in before and after studies (direct observation of vehicle use) can be used to evaluate changes in activity associated with corridor improvements. Application of marketing data as a supplement to license plate data (to provide households travel related attributes) is tested for the first time in this study. For this study, a large set of marketing data were purchased, where the total cost of marketing data acquisition and license plate data collection was approximately 20 cents per household.

Although the marketing data are available at the household level, and claimed by companies to be accurate, the data accuracy and precision are ambiguous and in need of rigorous examination. This section investigates the accuracy and inclusiveness of marketing data in comparison with self-reported data at both aggregated and disaggregated levels.

The purpose of this section is to examine the quality marketing data as a supplement for expanding the application of recent innovations in travel information sources such as cell phone data, electronic toll collectors, and ALPRs that lack the demographic component, targeting disadvantaged groups, and generating synthetic population. Each essential household attribute in transportation planning studies such as household size, number/presence of children, vehicle ownership, income, ethnicity, residential ownership and dwelling type has been compared across the sources in terms of accuracy and inclusiveness using statistical hypothesis testing.

The Atlanta household travel survey, which collected household level data for more than 10,000 household in 2011 is assumed to be a reasonable representative of "ground truth" in the comparative analyses. It is noteworthy that even self-reported survey data may not be 100% accurate. For example, people may underestimate their income level. However, self-reported survey data provides acceptable level of accuracy for the comparative purpose of this study.

Aggregated data have been obtained from American Community Survey at the Bock Group level (5 year summary file 2007-2011). Marketing data for the same 10,000 households have also been acquired. In addition, vehicle registration data for the 10,000 households are included in the comparison (only for use in assessing household vehicle ownership).

Selected socioeconomic variables applicable throughout transportation studies are listed in Table 1. The availability and coverage of the selected variables are tabulated across different data sources. As it is explained in the table legend, "A" and "D" refer back to aggregated versus disaggregated status of the variable. While the percentages illustrate the coverage of household specific values, in parenthesis percentages refer back to the availability of certain variable both from household specific as well as household inferred sources inferred by market agency using data imputation techniques.

Table 1 Socioeconomic Variables Availability and Coverage across Different Data Sources (N=10,000)

	ACS 5 year Summary File (Block Group)	Atlanta Household Travel Survey	Marketing Data	Vehicle Registration Database
Time Frame	2007 to 2011	Feb-Oct 2011	March 2013	March 2013
HH Size	A/100%	D/100%	D/89% (99%)	NA
Num of Children	A/100%	D/100%	D/89% (99%)	NA
Num of Adults	A/100%	D/100%	D/90% (99%)	NA
Num of Students	NA	D/100%	NA	NA
Num of Workers	A/100%	D/100%	NA	NA
Num of License Holders	NA	D/100%	NA	NA
Income	A/100%	D/100%	D/36% (99%)	NA
Vehicle Ownership	A/100%	D/100%	A/100%	D/100%
Ethnicity	A/100%	D/100%	D/98%	NA
Head of HH Age	NA	NA	D/82% (99%)	NA
Individuals' Age/Gender/Position	NA	NA	D/90%	NA
Family Composition	NA	D/100%	D/97%	NA
Education	A/100%	NA	D/16.7% (99%)	NA
Occupation	NA	NA	D/38% (99%)	NA
Residence Type	A/100%	D/100%	D/98%	NA
Residence Ownership	A/100%	D/100%	D/73% (99%)	NA

A= Aggregated, D=Disaggregated, NA=Not Available, ( )=Including Inferred Cases

The definition of quality data transcends typical "quality measures," such as "zero defects" or 100% accuracy. Rather, the definition of quality data dynamically hinges on how useful the information is to meeting specific research needs. Because data quality considerations vary by research purpose, the outcome often largely depends on how well the available data accomplish the study mission. For example, discrete choice modeling requires accurate data at the individual level, whereas a travel demand modeler needs to know more about travel behavior of various geographic and demographic cohorts.

Traditional considerations for data effectiveness include, but are not limited to, the availability, accuracy, timeliness (i.e., the data are current), and coverage of desired data elements. In the case of marketing data, third-party providers offer dictionaries with hundreds of data elements. Realistically, only a portion of the available variables are relevant for a particular transportation study. A different consideration to evaluate a data quality is about the usefulness of the data source relative to the research goals and objectives. In other words, predictive power and study value of the data should also be considered relative to any research need. The following sections describe assessment methods applied to evaluate different attributes of data quality.

4.2.1. Data Availability

Availability of various data elements provided by marketing companies and publicly available sources differ significantly. The first point of difference is whether the data are available at the disaggregated versus aggregated level. The aggregated variables, provided by ACS at block group level, are more applicable for area level studies. Specifically, the scale variables such as household size and income aggregated values are provided as means or medians, whereas non-scale (nominal or ordinal) variables such as age group, ethnicity, education level, residence type and ownership are provided as ratios (for example ratio of African-American households). Hence, while aggregated scale variables can be compared to household/individual level data numerically, the comparison of aggregated non-scale variables to household/individual level is practically not possible. Table 1 illustrates the availability of socioeconomic variables across the data sources ("NA" illustrates the variable is not available from that data source).

For household travel surveys, data availability is directly dependent on the survey questions and response rate. Typically, increasing length of surveys decreases response rate. Therefore, there should be a balance between length of survey, number of surveys and cost to achieve certain response rate for needed data elements. For example, education level, occupation, age and individual level attributes were not asked in the Atlanta household travel survey. However, the main advantage of self-reported surveys in transportation studies is the possibility of including data elements related to travel behavior either in the form of travel diary surveys or stated preference surveys that are not available from other sources.

Generally, almost any detailed information about households and individuals in the households are available in marketing data sources. However, increasing the number of data elements increases the cost of purchase. Additionally, while having access to detailed household information could increase the predictive power of our studies, increasing the number of households could also increase the statistical impact of the study. Therefore a balanced decision with respect to purchasing marketing data should take into account data elements availability, number of households, and cost of purchase, all at the same time. In this study, the only household level socioeconomic variable that could not be obtained from marketing data was vehicle ownership which was acquired from registration database instead. Lastly, not having access to some important travel behavior variables such as primary commute mode, number of daily trips, trip origins, destinations and purpose, and etc. is the main disadvantage of marketing data in terms of availability. A combination of household level marketing data with GPS tracking cell-phone data could provide a relatively complete household level travel survey data applicable for various transportation study purposes.

4.2.2. Data Coverage

Coverage is a major aspect of advertising marketing data and refers to a ratio of the availability of each variable across all the households in the study. The importance of coverage is mainly for marketing data whereas for self-reported aggregated or disaggregated data the coverage is close to 100% (including only respondents who answer all questions). It is possible that people ignore certain questions such as income; however, in this study all the available data elements across self-reported aggregated (ACS 5-year summary file) and disaggregated (Atlanta Household Travel Survey) have 100% coverage.

On the contrary, the coverage percentages across different data elements of the marketing data vary significantly. Marketing agencies try to utilize inferential techniques to infill missing variables and increase their coverage ratios. Looking back to marketing data column in Table 1, the ratios outside the parenthesis indicate real coverage values while the ratios inside the parenthesis indicate inferred coverage ratios. The real coverage ratios are not similar across the variables; very low for education and occupation (16% and 38% respectively), 36% for income, around 90% for household size related variables and 98% for ethnicity; the inferred ratios are entirely about 99%. This fact implies that marketing agencies attempt to fill out all the missing variables rather than leaving them empty. This means that the accuracy of inferred versus real attributes is also an issue. Leaving out the inferred attributes decreases the sample size significantly while keeping them in the dataset may have a negative impact on analysis if the data are not accurate. The accuracy section also evaluates the effect of including versus excluding the inferred attributes on data quality.

4.2.3. Data Timeliness

The Atlanta household travel survey was conducted in 2011 (February to October) and the marketing data were obtained in March 2013 for the same households. There is almost a two-year time lag between the two datasets. It is important to use recent data in analytical work using demographic variables, especially where such demographics change over time. Albeit, attributes will have different sensitivity to timeliness; therefore, presence of time lags between data sources would certainly affect the accuracy evaluation results. Thus, knowing to what extent time can negatively impact the accuracy assessment results is very important. To understand the impact of the two-year time difference, this section analyzed the Commute Atlanta travel survey data for household demographic stability.

The Commute Atlanta longitudinal dataset was used to assess the effect of time differences between data sources on data accuracy. The Commute Atlanta dataset includes detailed demographic information and vehicle activity data of 95 Atlanta households for 21 months from October, 2004 to June, 2006. The Commute Atlanta data provide a rare opportunity to examine the stability of demographic characteristics within households over multiple years. Commute Atlanta noted significant demographic instability in the participating households; 67 out of the 95 households underwent one or more significant demographic changes in at least one of six categories, including home location, work status, household structure, income, schools attended, and vehicle ownership (Xu et al., 2009).

Because the travel survey data, marketing data, and vehicle registration data are not obtained at the same time, the discrepancies among the data sources could in part come from longitudinal changes within households. Further analysis of the Commute Atlanta data was conducted to shed light on the magnitude of within-household demographic changes. Figure 10 presents the percentage of the 95 Commute Atlanta households that experienced changes in about two years (21 months). The demographic characteristics examined in Figure 10 mirror the key variables analyzed in this study. Changes in home location, household size, number of adults, number of children, and number of vehicles are self-explanatory. Changes in income are counted when a household's income shifted among the five income categories defined in the Commute Atlanta study: less than $29,999; $30,000-$49,999; $50,000-$74,999; $75,000-$99,999; and $100,000 or more.

Observations from the Commute Atlanta data indicate that 12% of all households moved within two years. In the context of the comparison between marketing data and travel survey data, the inference from the Commute Atlanta study is 12% of the households pulled from the same address may not be the same households. Number of vehicles tends to be a category with the highest change rate (31%), followed by income (18%). And, 14% of the 95 households experienced changes in household size, among them 11% had changes in number of adults and 4% had changes in number of children. Note that a household could experience a simultaneous change in number of adults and number of children (1%). The authors would like to point out that the Commute Atlanta sample only included households with at least one vehicle and is slightly biased towards higher income households, and arguably, higher-income households may have higher demographic stability. The households that participate in a travel survey may experience even more within-household demographic changes over time.

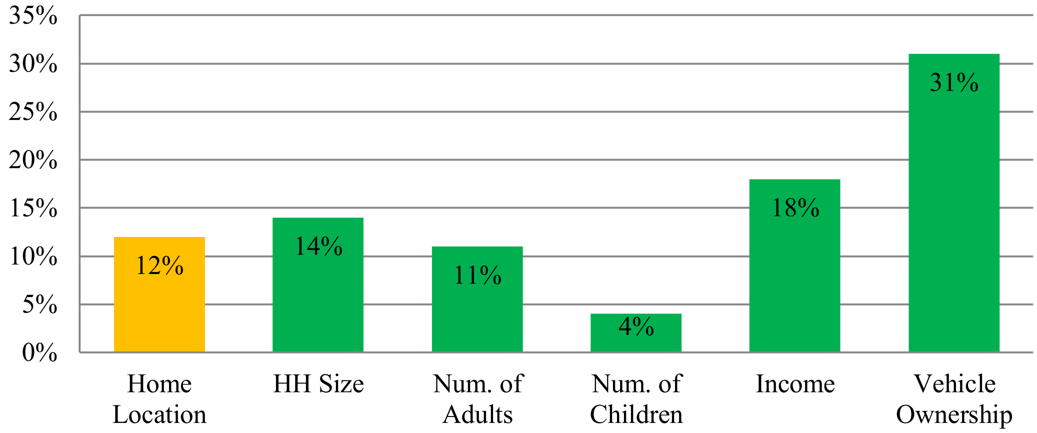


Figure 10 Household Demographic Changes Observed in Commute Atlanta Study (N=95)

It is important to note that for each variable within household, changes in addition to home location should be considered simultaneously. For example, we would expect that vehicle ownership differs by about 43% (12% + 31%) between the data sources with two-year time lag. The high sensitivity of socioeconomic variables to time highlights the need for a data collection method which enables repeated data collection over short periods of time, to keep the transportation planning and analysis current.

4.2.4. Accuracy Assessment

Desired level of accuracy in a dataset is very dependent on the study purpose. Therefore different methods of accuracy assessment have been applied here to present a comprehensive comparison between the data sources. To assess the accuracy of any data, a reference data is needed for comparison. In this case, disaggregated self-reported data (Atlanta household travel survey) is expected to be closer to the truth compared to aggregated and marketing data. However, certainly some participants do not report complete and accurate survey data. The statistical methods of comparison between the datasets are explained in the following sections. In any case that a reference category was required, Atlanta household travel survey has been assumed and conclusions should be tempered accordingly.

4.2.4.1. Exact Match Rate

Exact match rate refers to a percent of cases that are exactly and pair-wise equal between the two datasets. Exact matches are not relevant for aggregated data (which are averaged across many households).

4.2.4.2. Correlation

Correlation is a measure of association between two variables X and Y and ranges between -1 to 1. A value of +1 or -1 indicates perfect association between X and Y, the plus sign occurring for positive relationship and the minus sign occurring for negative relationship and zero would conclude that the variables are uncorrelated (Hayter, 2011).

The numerical variables in this study are either count variables (number of adults, children, and vehicles) or they are ordinal variables (income categories) without normal distributions (sig=0.000). Therefore, Pearson correlation coefficient which is applicable for normal continuous variables is not statistically appropriate.

A nonparametric measure of association between two variables, given by the Spearman's rho correlation coefficient, has been used in place of parametric Pearson correlation coefficient. Although in most cases X and Y are two separate variables, such as vehicle ownership and number of daily trips, in this case X and Y are two identical variables from different data sources. While exact match rate illustrates how well individual data agree, correlation coefficients shows how much correlated are two data sources for each variable.

4.2.4.3. Comparing Means/Medians

While the main accuracy evaluation should consider data quality at the case level, the central estimates of each numerical variable across data sources have also been evaluated. The underlying reason for this evaluation is to see whether any data source is biased toward higher or lower values. The means and standard deviations have been calculated in addition to means and standard deviation of differences between datasets which will be discussed in the results section. Tests of equality of means (paired Student's t-test) as well as equality of medians (Wilcoxon signed-rank test) have also been conducted. The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used when comparing two paired samples to assess whether their population medians differ (i.e. it is a paired difference test). It can be used as an alternative to the paired Student's t-test, when the population cannot be assumed to be normally distributed which is the case here (Sheskin, 2003).

4.2.4.4. Comparing Distributions (Statistical Tests)

After evaluating the accuracy at case level, as well as central level, the distribution of variables will also be compared. Although distributions comparison does not provide information about case level accuracy, the test indicates whether any data source has higher tendency toward specific sub-population groups. Moreover, if two data sources do not provide 100% exact match rate while providing similar distribution, one might assume that the errors are randomly distributed. In this case, although each case may not be 100% accurate, groups of cases could indicate that the data provide a reliable demographic profile in the region.

The distribution comparison has only been conducted between marketing and disaggregated self-reported data since the data types are the same; whereas in aggregated data the aggregate functions convert the ordinal form of the data to scale variable.

Several methods of comparison have been utilized here. First, distributions have been compared adopting non-parametric statistics for categorical variables (ordinal/count variables). For each variable, a crosstab table has summarized the frequency of all the cases where each axis represents data sources categories. The McNemar test is a simple way to test marginal homogeneity in K×K tables (Sheskin, 2004). Marginal homogeneity implies that row totals are equal to the corresponding column totals. Whereas this well-known method tests row/column homogeneity with respect to each individual category, the Stuart-Maxwell test tests marginal homogeneity for all categories simultaneously (Fleiss and Everitt, 1971).

The second test applied here is the Chi-Square goodness of fit test. The test assesses how "close" one distribution is to an assumed true distribution. This test is commonly used to test association of variables in two-way tables, where the assumed model of independence is evaluated against the observed data (Hayter, 2011). Both tests have been conducted and the results will be discussed in the next section. Lastly, the distributions have been presented in frequency graphs to facilitate the visual comparisons of the distributions.

4.2.5. Results

The accuracy evaluation is only applicable to those variables for which at least two data sources provide coverage in the same format. For example, there is no way to evaluate the accuracy of education, occupation, age, and individual’s attributes because marketing data source provides these variables at household/individual, while ACS provides the ratios in each category across block groups, and survey data didn't collect these attributes. For the variables for which a comparison is possible in practice, accuracy assessment results across different sources using appropriate statistical methods will be presented and discussed in this section.

The first quality assessment measures listed in Table 2 are the exact match rates. The exact match rate ratios indicate the percent of cases in disaggregated self-reported data (Atlanta Household Travel Survey) that have the same exact value in marketing dataset. For vehicle ownership, instead of marketing data, registration data have been used because marketing data do not provide household level vehicle ownership. The percentages include the inferred cases wherever is available these percentages vary across different attributes' categories. The total match rate percentages without inferred values are also tabulated in Table 2. Surprisingly, the inferred values are very good, considering the close total percent match rate with and without inferred cases. Therefore, for this set of marketing data inferential values do not appear to decrease the accuracy to such an extent that one would specifically exclude inferred data from analysis. It should be noted that there are large differences between numbers of cases at each category of the attributes. The frequencies have not tabulated in the same table to avoid over-populating table cells. Meantime, the frequencies will be presented in the form of frequency bar charts later in this section.

The exact match rate for overall household size for non-inferred and non-inferred plus inferred is only about 33%, which is not as high as one would like to see. For household size, and number of adults to have the value of "1" the match rate is 66% and

61% respectively and decreases as the household size and number of adults increases. The same trend has been observed for number of children with match rate of 81% for zero children. Thus, employment of marketing data makes more sense when the analyst main concern is presence of children or low occupant household rather than the exact number of household members. In general, number of children with 60% total match rate has better accuracy in relation to number of adults with 48% and household size with 32%. Considering total match rates, it is more reasonable to treat household size as two separate variables of children and adults in future analysis of marketing data to avoid increasing inaccuracy by summing the errors. Meanwhile, 14% change in average household size and 12% change in home location over two years Commute Atlanta study period could explain a significant portion of the difference comparing two data sources.

Income categories total match rate between marketing and survey data, based on six categories provided in the Table 2, is 34%. The match rate increases to 41% by removing inferred cases which represents 62% of total cases. The match rate is also very dependent to number of categories of comparison meaning higher match rate for larger income categories. For example, if three categories instead of six have been used, the match rate increases to 58% with inferred data and 64% without inferred data. The match rates are very close across the categories implying close to normal distribution of error. Similar to household size, removing the inferred cases omit 62% of the cases while improve the quality of the analysis slightly.

Regarding vehicle ownership, the match rate was lower than what was expected since both of the data sources (survey and registration) are assumed to be very close to reality. Yet, match rates are affected by the fact that many registered vehicles are not being used at their place of registration for multiple reasons. For example, children's registered their vehicle at their parents' house for insurance purposes (Granell, 2002). The highest match rate for vehicle availability is for households with one and two vehicles which are 62% and 52% respectively. In general, vehicle availability match rate is 48%. Meanwhile, 31% change in average vehicle ownership and 12% change in home location over two years Commute Atlanta study period could explain a significant portion of the difference comparing two data sources.

Table 2 Exact Match Rate Percents across Different Attributes' Values between Marketing Data and Travel Survey Data (N=10,000)

Variables	Variables Categories						Total (Including Inferred Cases)	Total (W/out Inferred Cases)
Household Size	0	1	2	3	4	5+		
Children and Adult	NA	66%	27%	22%	15%	10%	32%	33%
Children	81%	24%	15%	15%	6%	3%	NA	60%
Adults	NA	61%	49%	19%	15%	6%	48%	48%
Income	< $29,999	$30,000 $49,999	$50,000 $74,999	$75,000$ 99,999	$100,000 $149,999	$150,000 <		
	41%	30%	32%	27%	35%	42%	34%	41%
Vehicle Ownership	0	1	2	3	4	5+		
	36%	62%	52%	35%	32%	22%	NA	48%
Ethnicity	White	African-American	Asian	Hispanic	Other			
	84%	56%	42%	64%	4%		NA	73%
Residential Type	Single family	Condo/ Apt	Mobile	Other				
	98%	13%	20%	0%			85%	85%
Residential Ownership	Own	Rent						
	95%	49%					89%	95%

Ethnicity has the highest match rate of 84% for white population with 40% to 60% for all other ethnicities. In total, ethnicity has 73% match rate which makes it a relatively accurate variable compared to the others. The fact that ethnicity does not change over time for a certain household has a role in having high quality measure. Hence, 12% change in home location over two years Commute Atlanta study period could explain a significant portion of the difference comparing two data sources.

Residential type and ownership have the highest match rates of 85% and 89% respectively. However, the match rate for condo residential type is significantly lower than single family households (13% vs. 98%) and the match rate of renters is significantly lower than the owners (49% vs. 95%). Thus, if the objective of a researcher is to find renters or condo occupants, the marketing data is not a good source while it is a more appropriate source for owners and single family units. The high quality of these variables could be explained by their stability at address level.

The rest of the accuracy assessment measures are tabulated in Table 3. The first measure is Spearman's rho correlation coefficient which in general is higher between self-reported disaggregated data and marketing data. This difference was expected because the aggregation removes all the cases' deviations from the center and creates much fewer low and high values in aggregated data compared to disaggregated data. In comparing the marketing and disaggregated self-reported data, number of adults has the lowest correlation of 0.29 compared to number of children and household size which are 0.47 and 0.41 respectively; whereas income and vehicle ownership both has the highest correlation of 0.6. Income is the only variable that aggregate data also shows high correlation of about 0.5 to both marketing and disaggregate self-reported data.

The next statistic is the mean and standard deviation for three data sources across all the available attributes. The first note is the considerably lower standard deviations for the aggregated data source which ties back to the aggregation form of the data (mean and standard deviation of the means/medians). Moreover, the aggregated data represent the highest average value for household size and the lowest average value for income and vehicle ownership. Marketing data represent the highest value of income. The registration data represents the highest value of vehicle ownership. Number of children highest value is reported by marketing data which is very close to the aggregate data. The highest value of number of adults is reported by aggregate data while the two marketing and self-report data reports very close values.

Table 3 Statistical Tests Results across Variables (N=10,000)

	Variables	Vehicle Ownership*	Income	Number of Children	Number of Adults	Household Size
Spearman's rho Correlation	SRA/SRD	0.295	0.499	0.18	0.17	0.176
	SRA/MRK	0.279	0.572	0.17	0.18	0.227
	SRD/MRK	0.601	0.609	0.476	0.297	0.405
Mean (SD)	SRA	1.84 (0.51)	$ 70,644 (32k)	0.75 (0.37)	2.01 (0.40)	2.73 (0.57)
	SRD	2.06 (1.07)	$ 82,516 (56k)	0.62 (1.00)	1.912 (0.79)	2.53 (1.32)
	MRK	2.26 (1.56)	$ 88,020 (55k)	0.77 (1.45)	1.91 (0.91)	2.62 (1.82)
Mean Difference (SD)	SRA-SRD	-0.22 (1.07)	$-11,872 (51k)	0.13 (1.03)	0.10 (0.85)	0.20 (1.3)
	SRA-MRK	-0.41 (1.5)	$-17,376 (46k)	-0.02 (1.44)	0.101 (0.96)	-0.12 (1.81)
	SRD-MRK	-0.19 (1.29)	$ -5,504 (37k)	-0.15 (1.3)	0.002 (1.05)	-0.08 (1.76)
Median Tests (Wilcoxon's) Sig	SRA / SRD	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	SRA / MRK	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	SRD / MRK	(0.000)	(0.000)	(0.000)	(0.199)	(0.000)
Distribution Tests (Maxwell's and Chi Square) Sig.	SRD/MRK	M(0.000), C (0.000)	M (0.000), C (0.000)	M (0.000), C (0.000)	M (0.508), C (0.000)	M (0.448), C (0.000)

SRD=Self-reported Disaggregated Data (Survey), SRA=Self-reported Aggregated Data (ACS), MRK=Marketing Data, M= Maxwell's Test, C=Chi-Square Test, * Marketing source for vehicle ownership is registration database.

Looking at the mean differences, marketing data and disaggregated self-reported data are closer to each other, and both are farther apart from aggregated data, especially for income. All the means are significantly different and the standard deviations of the differences are also presented in the parenthesis indicating that 68% of the cases are less than one unit of standard deviation different. Using income as an example, for 68% of the cases, differences between marketing and disaggregated data are in the range of $-5,504 ± 37,000 and for the remaining cases the differences are beyond this amount. Small differences in mean values imply that distributions have very close central values while large differences illustrate the extent of discrepancies between the data sources.

Regarding vehicle ownership, more than one standard deviation of difference between registration and self-reported data is again ties back to the large discrepancy between numbers of vehicles registered at each household versus number of vehicle actually under use at each household (Granell, 2002).

In addition to test of the means, test of the medians have also been conducted (Wilcoxon's test) which illustrates significant different medians for all the attributes except number of adults with 95% confidence.

Lastly, Chi-Square goodness of fit tests indicate significantly different distributions for all the attributes, while Maxwell's indicates not significantly different distributions for number of adults and household size with 95% confidence.

Given the noted statistically significant differences in the variable distributions, Figure 11 illustrates the distributions of attributes from marketing, disaggregated self-reported and registration data sources to help identify the deviations. The reason that aggregate source distributions have not been presented is the absolutely different form of distribution (continuous and very close to normal) because of the aggregating functions. Once again, the distribution comparison provide no information about individual cases' accuracy but provide insight into the presence of any bias associated with marketing data compared to self-reported disaggregated data.

The household size attributes show large differences in the 1 and 2 values, while distributions for number of children are similar across the self-reported survey data and marketing data. The difference in number of adult (1 vs. 2) appears to dominate the impact on household size low match rate. Accordingly, the main difference in distributions of household size is much more cases with "1" and fewer cases with "2" and more household sizes. Therefore, the marketing data applied in this study may over-represent households with single adults with or without children, and under-represent households with couples with or without children. Considering the potential bias, upcoming socioeconomic analyses and modeling sections treat household size as two separate variables of children and adults, to keep the potential bias of the number of adults separate from number of children.

As a result, if we treat marketing data household size with following categories: "1 or 2", "3", and "4+", the distributions would become very similar. Unfortunately, in travel behavior, households with one person behave differently than two-person households in transportation decision making. So, the feasibility of this new classification method depend upon research purpose and the importance of knowing if a household has one versus two members.

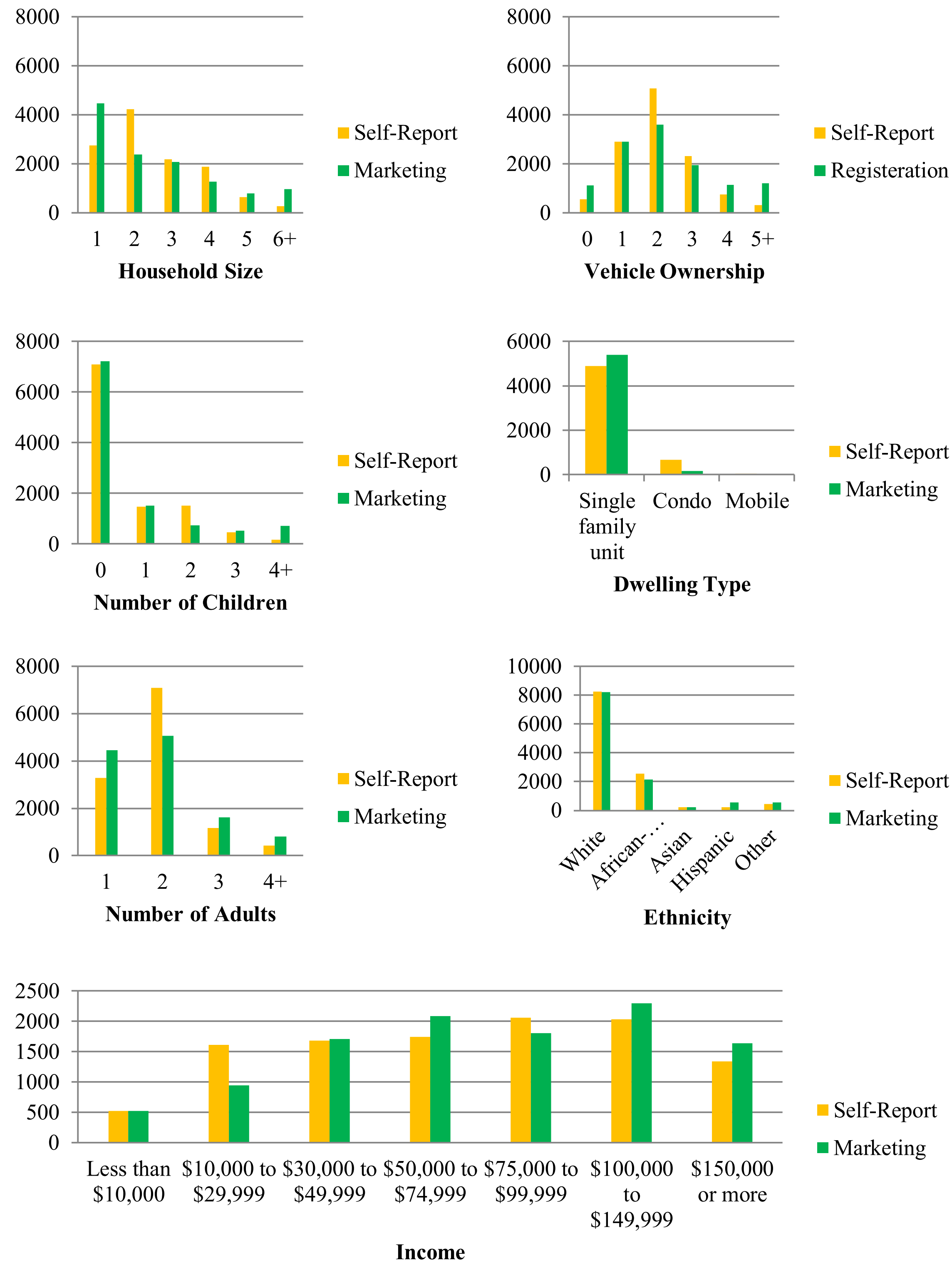


Figure 11 Distributions of Disaggregated Self-reported Data (survey) and Marketing Data (N=10,000)

Regarding vehicle ownership, self-reported data present more households with two and fewer with 4 and 5+ vehicles compared to the registration database. The underlying reason for this discrepancy may be the fact that people only report the vehicles that they use rather than vehicles that are registered at their home (Granell, 2002). Because vehicle ownership is a critical variable in many transportation studies such as travel demand modeling and air quality, this dissimilarity in number of vehicles between registration and self-reported data could be very important and needs to be considered; specifically, when both registration and self-reported data are treated as the truth in many studies.

Regarding dwelling type, as it was discussed before, marketing data tend to report more single family unit compared to condo/apartments; however, the difference is not very significant. Regarding ethnicity, the distributions are very close, which makes marketing data a very reliable source of ethnical/racial studies in the area of environmental justice and social sustainability research.

Lastly, regarding income, marketing data tend to report higher household income in contrast with self-reported data. This issue may tie back to the fact that one main source of marketing data is credit report data, in which people may tend to report higher income for themselves to attain larger loans or credit card limits. Nevertheless, the income distributions are not practically that much different which probably make marketing data an acceptable source for economic analysis; particularly, when the income dissimilarity between groups of population is the main point of study rather than the exact amount of income.

4.2.6. Discussion

The accuracy of marketing data is very dependent upon time. Marketing agencies claim to update their entire database in less than a quarter. Some attributes such as vehicle ownership and income are more sensitive to time while household size related variables are less sensitive (based on longitudinal Commute Atlanta study). Some attributes such as ethnicity is not sensitive to time at all (except when whole households move).

Additionally, the address dependency of marketing data, in license plate-based studies (household location from registration data base) creates another source of error of home location (people do not live at where they register their vehicles or they move between the time the data is collected in the field and the socioeconomic data is acquired form marketing sources). About, 33% of the total vehicles were not registered in the same place that they start their daily trip (Nelson, et al., 2010). Based on GIS spatial tools, more than 96% of the registered vehicle addresses do fall within the corridor commutershed (Khoeini et al., 2012).

Analysts should decide about the application of marketing data based on the objective of the study. Marketing data source could supplement travel data (collected via ALPRs, tag readers and cell phones) in terms of socioeconomic attributes, when the time lag between obtaining trip data and marketing data acquisition is not very long (preferably less than a year); additionally, when the trends of differences between groups are more the focus of attention rather than each individual household travel behavior. Particularly, in case of limitation in time, labor and money resources and urgent need for conducting a preliminary study, application of marketing data is recommended.

Furthermore, marketing data can assist household travel surveys by identifying targeted market households (such as disadvantaged, disabled, and etc.), generating synthetic population as well as data validation, imputation and augmentation.

CHAPTER 5

PRELIMINARY SOCIOECONOMIC ANALYSES

This chapter summarizes the analytical methodologies and results of four preliminary analyses. The first section investigates the sensitivity of users block group level socioeconomic attributes to license plate data collection time of day, day of week, and location of data collection. The second section examines vehicle value as a proxy for household income for socioeconomic analysis relative to congestion pricing, which is a case study on Atlanta I-85 HOT (High Occupancy Toll) corridor. The third section explores the impact of conversion on HOV(High Occupancy Vehicle) lane frequent users. And, the last section utilizes block group level attributes for socioeconomic analysis and modeling of the I-85 HOV and HOT lanes usage.

5.1. License Plate Data Collection Sensitivity Analysis[1]

Although license plate based demographic analysis for Atlanta corridors has been performed before, these analyses have not investigated the sensitivity of the demographic results to data collection variables such as time-of-day (e.g. morning vs. evening commute periods), day of the week, and different locations for data collection (Granell, 2002; Nelson and Zuyeva, 2010; Nelson et al., 2008). One reason for the lack of these prior analyses is that accurate sensitivity analysis requires a very large sample size which is labor-intensive, expensive and time-consuming. The large dataset (with more than

[1] This section has been published: S. Khoeini, R. Guensler, M. Rodgers, V. Elango. "Sensitivity of commuters' demographic characteristics to license plate data collection specifications: A case study for HOV-to-HOT project in I-85 corridor, Atlanta, GA"; Transportation Research Record: Journal of the Transportation Research Board, National Academy of Sciences; No. 2308, pp. 37-46, 2012.

189,000 license plate observations) available in this study makes reliable sensitivity analysis possible.

The analysis in this section uses only the fall 2010 license plate data for which processing was completed in time for analysis. In fall 2011, although two sessions during each four days of the week (Mon, Tues, Wed, and Thurs) for all 5 locations were planned, some sessions were cancelled due to weather conditions or scheduling conflicts. The vehicle volumes are not equal for all the sessions and sites; therefore, the number of collected license plate in each session is not equal. However, this difference is statistically negligible due to the large sample sizes available in each category (more than 20,000 cases).

Using more than 189,000 license plates collected in fall 2010, this section assess whether there are any significant differences between the demographic characteristics of different commuters groups divided based on time (AM peak, PM peak), day of week (Mon, Tues, Wed, Thurs) and location (Beaver Ruin Road (BRR), Chamblee Tucker Road (CTR), Pleasant Hill Road (PHR), Jimmy Carter Blvd (JCB), and Old Peachtree Road (OPR)) of data collection.

Three main hypotheses are tested:

- Whether morning and afternoon commuters exhibit different demographic characteristics.
- Whether Monday, Tuesday, Wednesday and Thursday commuters exhibit different demographic characteristics.

- Whether different data collection sites exhibit different demographic characteristics.

Two main demographic characteristics have been considered for this analysis. One is household size, from Census 2010 at the Census block level and the second is household median income. Because Census 2010 did not collect any income data, the American Community Survey (ACS) 5-year estimate (2005 to 2009) has been used to represent household median income at the block group level (highest available resolution from Census data). Table 4 shows percentile of population in GA and HOV corridor commuters of I-85 with income less than figures presented in the table. For instance, household income for 50% of GA residents is less than $51,250 while 50% of I-85 HOV commuters have income less than $69,699 which is much higher.

Table 4 Comparing GA Population Income and I-85 HOV Commuters

Percentile	GA Population Household Income	I-85 Commuters Household Income
10%	$ 28,580	$ 39,595
20%	$ 35,455	$ 47,903
30%	$ 40,862	$ 55,438
40%	$ 45,850	$ 63,942
50%	$ 51,250	$ 69,699
60%	$ 57,774	$ 74,578
70%	$ 65,625	$ 80,213
80%	$ 74,102	$ 87,857
90%	$ 91,058	$ 93,125

5.1.1. Sensitivity of Demographic Results by Time of Day (AM, PM Peak)

This section assesses the sensitivity of demographic data to the time of data collection. This sensitivity is not just about the time but also direction of commuters. During AM peak hours, only southbound (toward CBD) commuter license plates were collected and during PM peak hours, only northbound (out of CBD) commuter license plates were collected.

In examining the effects of time of the day, data from JCB were examined. JCB is located in the center of the corridor and therefore, include the commuters from I-285 and US-316, which are the two major highways intersecting I-85 along the HOT corridor (Figure 2).

Two subsets of data (Site: JCB, Time: AM and PM) were selected in SPSS. Test of normality were rejected with a p-value less than 0.001. Therefore, nonparametric tests were used to compare household size and household income between two groups. The Mann and Whitney test which is the most commonly used alternative to the independent-samples t-test for normally distributed data, was rejected with p-value less than 0.001 for both household size and income. Given the very large sample size, this is not surprising. Figure 12 shows the box and whisker plot for different JCB, AM and PM groups.

Note that the mean and the variability between the groups are very similar. The difference between the household income mean is only $378 and the difference between the household size is only 0.024. From a practical perspective, these differences are not meaningful, given the large average value. While, most of the morning trips are toward work zones, some of the afternoon trips include purposes other than work, such as shopping, leisure, etc. Hence, the afternoon traffic may contain a larger percentage of lower income persons from smaller households heading out of the city for other activities. Further investigations are provided in Chapter 8 where AM and PM commutersheds are compared.

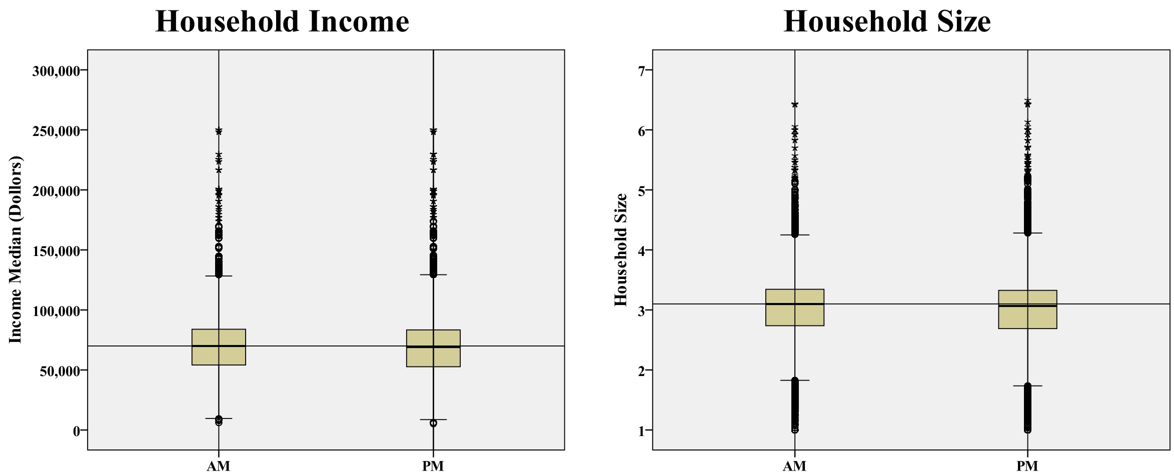


Figure 12 Box and Whisker Plot for Different Times of Data at JCB (AM Number of Records: 26,656, PM Number of Records: 22,553)

5.1.2. Sensitivity to day of the week (Mon, Tues, Wed, Thurs)

Because HOT implementation is designed to provide reliable travel times for HOT lane users during peak-hour congestion, weekends have not been considered in the data collection. Monday, Tuesday, Wednesday and Thursday are four days of the week that have been considered. However, because of some missed sessions of data collection, not all days of the week have same number of records. A random sample of 120,000 records (30,000 per each day) has been selected for the statistical analysis. Figure 13 shows the box and whisker plots of household income and household size associated to different days of data collection. The average and standard deviation for household income and household size is not significantly different across different days of the week. However, to support a more comprehensive comparison, the distributions have to be compared using statistical tests.

Test of normality of household income and household size for four subset of data associated to four days of the week were rejected. Therefore, nonparametric test were used to compare the distributions. The first test is between Mon, Tues and Wed in JCB afternoon session. The Kruskal-Wallis test, which is a nonparametric alternative to one-way analysis of variance (ANOVA), cannot reject the null hypothesis for the household income with p-value equal to 0.669 and cannot reject the null hypothesis for household size with p-value equal to 0.099. Therefore, distributions of first three days of the week (Mon, Tues and Wed) are not deemed to be statistically significantly different. However, the same test for CTR between Mon, Wed and Thurs, fails to reject the null hypothesis for both household size and income. This means that Thursday may have a different distribution compared to the first three days of the week. However, the change from JCB to CTR should also be considered. CTR which is the first station in the corridor has more variability compared to JCB which is in the middle of the corridor.

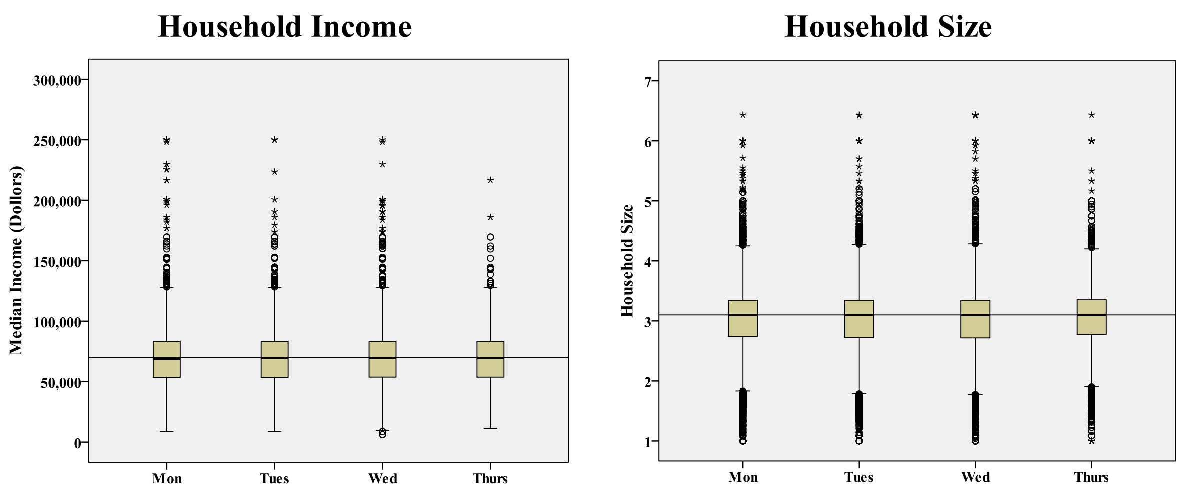


Figure 13 Box and Whisker Plot for Different Days of Data Collection at BRR (Number of records: 30,000 per day)

To compare Thursday to other days of the week more precisely, same Kruskal-Wallis test has been used to test the difference between three days of the week (Tues, Wed, Thurs) at PHR. The test rejects the null hypothesis for household income while it fails to reject the household size. This result indicates the likely significant effect of data collection site on demographic attributes of the users.

5.1.3. Sensitivity to location of data collection

Because the number of available cases is different in each site, 15,000 cases were selected from each site at random for the statistical analysis. Figure 14 shows the box and whisker plot of the data. Average household size and household income as well as variability appear very similar across different sites. The variability of income at CTR is larger than the other locations. The main reason for this is likely to be the presence of the I-285 to I-85 interchange immediately downstream from the CTR location (Figure 4). Vehicles identified at CTR merge with vehicles arriving from the I-285 and together compose the traffic observed downstream at the JCB, BRR, and PHR sites. Hence the distributions of income and HH size may be significantly different at CTR when compared to the other locations along the corridor, depending upon the correlation of the household demographics and employment type and whether those jobs are located in the central city or along the I-285 suburbs.

The maximum difference in household size mean between the five sites is less than 0.1 persons per household, which is negligible. However, the maximum difference in income is $3000 (5% of the average mean), which is not likely to be considered negligible for most planning purposes. Differences are also noted for the data at OPR, likely because the I-85 outbound traffic splits at the SR316 interchange, with part of the traffic exiting on SR316 and part of the traffic remaining on the I-85. CTR and OPR, which are located at the two ends of the corridor, yield the highest and lowest income values. The other three sites lie somewhere in between and do not differ from each other in practical terms.

The five data subsets associated with the different sites are not normally distributed, so non-parametric tests were used to compare the distribution of household size and household income across the five sites. The Kruskal-Wallis test, which is a nonparametric alternative to one-way analysis of variance, rejects the null hypothesis with p-value less than 0.001 for household size and household income across five different sites. Again, this is not surprising given the very large sample sizes involved.

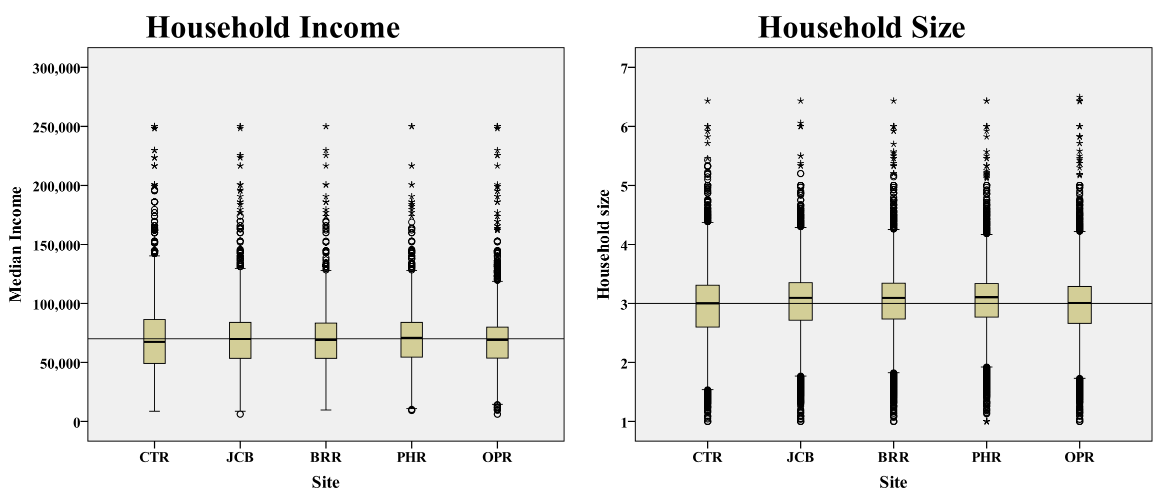


Figure 14 Box and Whisker Plot for Different Sites of Data Collection (Number of records: 15,000 per site)

As discussed earlier, CTR is located inside I-285 and exhibits less variability compared to the other locations. OPR, which is the last location for data collection after the SR316 split, also appears to have less variability compared to the others. The income is significantly lower at OPR compared to the other locations. One reason is that from JCB to BRR and BRR to PHR, many high-income residential areas are accessed by vehicles departing the freeway corridor. However, once the end of the Atlanta

Metropolitan Area is reached, the outlying residential areas consist of lower income households. Hence, vehicles that remain on the system after the SR316 exit represent households with a lower average income.

5.1.4. Reducing Data Collection Efforts

Because license plate data collection is time-consuming and labor-intensive, reducing the number of data collection sessions based on statistical analysis is desired. If the same conclusions can be reached through statistical analyses using smaller data sets, then fewer data will need to be collected. The CTR and OPR data collections sites located at the two end of the corridor exhibit the highest difference in demographic variables compared to other locations and are important to retain in future studies. However, it may be possible to exclude one or two locations for data collection out of remaining three sites (JCB, BRR, and PHR).

Three paired sets of data were selected and Mann-Whitney test were executed. The JCB and PHR sites are not adjacent to each other and BRR site resides between them. Therefore, JCB and PHR do not have the same statistical distribution in household size and income. However, JCB and BRR are fairly close together and do not appear to have a statistically different distribution of income (p-value = 0.49) but do have a slightly different distribution for household size (p-value = 0.022). The other two adjacent sites (BRR and PHR) appear to have a same distribution for household size (p-value = 0.213) but do appear to have a different distribution for income (p-value < 0.001).

In conclusion, it may be possible to eliminate the BRR site from future data collection without significantly impacting observed demographic distributions. Because the team was conducting a before-and-after study, it was important to consider whether dropping the BRR site from future data collection would have an impact on observing changes in the on road fleet. That is, if the demographic profiles of drivers entering between the site eliminated and the sites retained are significantly different than the drivers already on the road, and their behavior with respect to HOT lane usage is different, it may be necessary to retain the site in the ongoing data collection efforts. However, because the effect of day of data collection was negligible especially for three middle sites, the team decided to decrease the number of data collection days to two days per week preferably Tuesday and Thursday for each middle site.

5.1.5. Summary

The sensitivity of demographic data to day of week appears to be much lower than time of day and site of data collection. As expected from travel demand theory, AM and PM distributions were different, likely due to differences in morning and evening trip purposes and commutersheds. While different sites exhibited different demographic characteristics, some differ more significantly, likely based upon the presence of intersecting corridors result in a change in the overall demographic profiles of freeway commuters. That is, the demographics revealed by the on road fleet change from inside the perimeter to outside the perimeter, and after the SR316 split as vehicles enter or leave the HOT corridor via I-285 or SR316.

Large sample sizes increase the chance of rejecting the null hypothesis. Yet, even though the differences are statistically significant, the actual mean and variance of demographic data for one session of data collection is very similar for all 28 sessions of data collection.

5.2. Vehicle Value Analysis[2]

While properly managed HOT lanes can provide a reliable travel option, the toll and occupancy requirements produce socioeconomic concerns regarding potential disproportionate distribution of any negative impacts across low income or other protected groups (Executive Order 12898, 1994). Thus far, self-reported survey data have been the main data sources for socioeconomic studies. Self-reported survey data provide valuable information about lane use, but these surveys are expensive and time-intensive. Furthermore, the significance, precision and accuracy of the findings depend upon sample size and representativeness of the households that choose to participate in the surveys. Collecting an unbiased and large enough sample of user characteristics across the wide variety of demographic variables of concern in transportation planning is very expensive and often not feasible for projects or studies.

Based upon the findings of previous studies, income is a key determinant for managed lane use, as well as equity and environmental justice assessment studies. Some studies used self-reported income data providing individual level accuracy for a small proportion of the users (Burris and Hannay, 2003). Other studies used license plate data matched to household address blockgroup providing area level aggregated data for relatively larger samples of users (Khoeini et al., 2012; Nelson et al., 2008). Despite the importance of income data, obtaining users' income at the household level via traditional survey methods is difficult and expensive.

---

[2] This section has been published: Khoeini, R. Guensler. "Using Vehicle Value In place of income For Pricing Economic Analysis: A case study on Atlanta I-85 HOT lane"; Research in Transportation Economics, Special issue on "road pricing in US", 2014 (In Press).

Traditionally, economists and market researchers have been interested in identifying the factors that affect consumers' car buying behaviors to estimate market share, and to that end they have developed various models of vehicle choice. These models are generally focused on vehicle attributes (such as operating and capital costs, horsepower, and fuel efficiency), household characteristics (such as number of household members, number of vehicles, and household income), principal driver characteristics (such as age, education, and income), and attitude and lifestyle attributes (Choo and Mokhtarian, 2004; Golob et al., 1997; Kitamura et al., 1999; Train, 1986). In the developed models, the income variable either directly, or in combination with other variables, was significant in predicting vehicle type choice; implying that having higher income households had higher probabilities of owning high value vehicles. Furthermore, Miller and Davis (2002) found a strong positive correlation between vehicle age and average household annual income.

Despite the positive relationship between vehicle value and income in the literature, vehicle value has not been used in studies as an indicator for a users' ability to pay tolls. This study proposed the employment of vehicle value as a surrogate for household income, particularly when not enough resources are available for collecting actual users' attributes.

Meanwhile, the advances in Automatic License Plates Reader (ALPR) technologies (Colberg, 2013) now facilitate license plate data collection and processing and have decreased the cost of ALPR data collection. Lastly, the widespread adoption of congestion pricing throughout the nation and the world raises the need for alternative low-cost socioeconomic impact assessment methods to ensure that low income and other protected groups are not disproportionately negatively impacted.

5.2.1. Methodology

To estimate vehicle value data, 341,589 license plates (representive of 101,264 unique households) were collected across all the lanes in spring and summer 2012 (six months after the HOV to HOT conversion) along five sites in the corridor, during morning and afternoon peak periods. The State motor vehicle registration database provides vehicle make, model and model year data for the collected license plates, for which online vehicle value estimation tools have been used to estimate vehicle value.

The costs of the applied method include license plate data collection, processing, and vehicle value estimation, commands a price of approximately 20 cents per household, which is roughly 0.1% of the cost of household travel surveys. However, the point that household travel surveys provides substantially more socioeconomic and travel related variables should not be disregarded.

To evaluate the users' willingness to pay tolls across the lanes, the value of the vehicles using the new HOT lane and adjacent general purpose lanes on Atlanta I-85 have been analyzed and compared. The assessment also looks at the distinction between HOT and general purpose lanes regarding the vehicle makes and models ranking, to investigate whether the popular term "Lexus Lane" can be reasonably applied to HOT lanes. After identifying the difference in vehicle value across the lanes, the analysis assesses the proportion of the difference that derives from the use of newer vehicles (higher vehicle model years) vs. the use of more expensive vehicle makes and models.

To support the assessment and verify the proposed methodology, the targeted market household income data have also been used to conduct a parallel analysis and examine differences in household annual income across the lanes.

This section has two main objectives: examine the application of vehicle value as a proxy for household income, and investigate the difference in vehicle value, vehicle type and users' income across different lanes along Atlanta I-85 HOT corridor. This study contributes to transportation practice by establishing an inexpensive method for potential socioeconomic impact assessment and to transportation policy by evaluating the Atlanta I-85 HOT lane impact on users' lane assignments based on their ability to pay toll (represented by household annual income and average vehicle value).

To obtain vehicle value estimates, several online sources were evaluated. Data accuracy and convenience of use were the two priorities in evaluating different data sources. Finally, TRUECar® (TRUECar, Inc.) and Kelly Blue Book (KBB) were selected. While KBB is probably the most recognized and is deemed a reliable source, its price checking process is not as convenient as TRUECar. TRUECar was used for all vehicles except the 2012 vehicle models, because TRUECar does not provide used car price for 2012 model. KBB was used for the 2012 vehicle models. The KBB used car price, rather than the dealership or invoice price was used for 2012 models so as to reduce the potential bias toward high value cars in the dataset.

In the TRUECar price checking process, the private seller price for vehicles in good condition with standard attributes was assumed for all vehicles. While different make, model and model year combinations in the observation data yield 21,734 unique combinations, 2,700 of these combinations represent more than 85% of the fleet. For practical reasons, the analyses proceeded with these 2,700 unique combinations. Four undergraduate students collected the data from the websites and a graduate student assessed the accuracy of the dataset.

By matching the vehicle make model year of the price dataset to the original license plate dataset, 85% of the observed plates were assigned a vehicle value. Because the entire price checking for this analysis was performed in the summer of 2012, it was not reasonable to use these prices for vehicles that were observed in 2010 or 2011. Research designed to assess the change in vehicle values from the HOV lane to the HOT lane was ongoing and required access to depreciation tables allowing the team to assign vehicle values to the vehicles at the time they were observed (i.e., 2010 and 2011).

Hence, this study is based on license plate data collected in spring and summer 2012 matched to the prices checked in summer 2012. Accordingly, the sample size of the study decreased to 341,589 license plates, representing 101,264 unique households.

5.2.2. HOT and GP Lanes Vehicle Value

Based upon the literature, which indicated that wealthier households are likely to use the HOT lanes more frequently, higher vehicle values are expected in the HOT lane compared to GP lanes. However, ability to pay toll (reflecting household financial status) cannot be considered the only reason which would cause a difference in vehicle value across the lanes. Because I-85 is a six lane expressway, vehicles along the left lanes particularly the most left lane (HOT lane) are potentially longer distance commuters. It is also possible that longer distance commuters choose more reliable vehicles compared to shorter distance commuters or drive faster regardless of their income.

Furthermore, corridor commutershed socio-spatial distribution could also have potential impact on users' lane choice. Figure 15 shows HOT corridor commutershed developed based on license plates and the census block groups for the associated registration data. Approximately 95% of corridor users reside within the ellipse (section 8.1.2.). The household annual income increases in moving from central district toward suburban areas up to the end of Atlanta metro area. Because the I-85 corridor is primarily used by commuters who work in the Atlanta central business district and live in Atlanta northeast residential areas, it is expected that people who live further from Atlanta (long distance commuters) will use the left lanes, and particularly the HOT lane, more frequently. Given the location of the higher-income residential areas (green block groups), a higher vehicle value in the left lanes will not be surprising. Further analysis and illustration of the corridor commutershed will be presented in chapter 8.

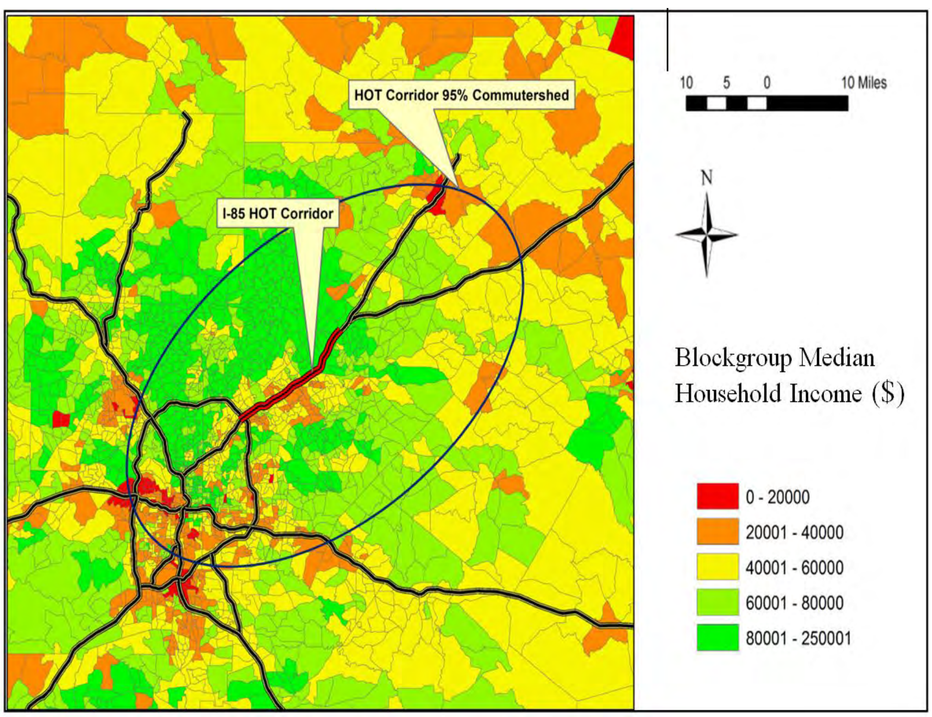


Figure 15 HOT Corridor Commutershed Household Annual Income

Figure 16 illustrates the 95% confidence intervals around the average vehicle value across the six lanes (the HOT lane and five GP lanes) and similarly Figure 17 illustrates the lanes average vehicle model year confidence intervals. The confidence intervals around the mean can be interpreted to mean that we are 95% confident that the average vehicle value falls somewhere within the confidence interval (but is equally likely to fall anywhere within the bounds). Hence, when confidence intervals do not overlap, we are reasonably confident that the means are different.

Figure 16 and Figure 17 illustrate the average vehicle value and model year confidence intervals across the lanes. As expected, HOT lane average vehicle value and average model year is significantly higher than the adjacent GP lanes. Moreover, vehicle values across the two GP lanes just beside HOT lane are also significantly higher than the three most right lanes. The similar trend could be observed for vehicle model year. While, a significant decreasing trend from HOT lane to GP3 lane has been observed in vehicle value and model year, no significant difference between the three GP lanes at the most right of the corridor was observed.

Table 5(a) and (b) show the detailed statistics for vehicle value and model year across the observed lanes, including the differences between HOT lane and GP lanes. Table 5(a) shows vehicle value statistics across all the lanes, while Table 5(b) aggregates general purpose lanes and adds the model year statistics as well. Because the distribution of vehicle value is similar to other monetary variables (i.e., has a long right-tailed distribution) the median is lower than the average value. However the difference between the medians is very close to the difference between the means.

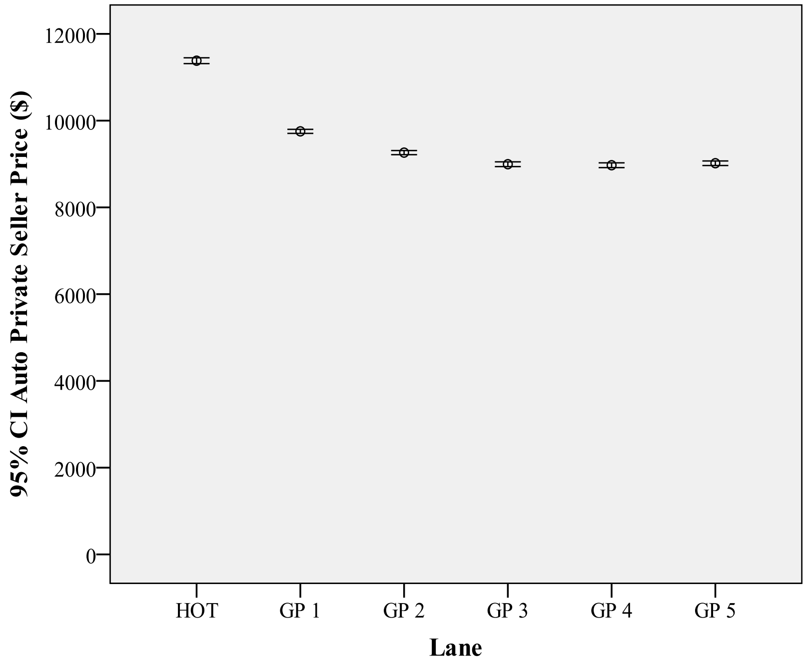


Figure 16 Confidence Interval for Private Seller Price across the Lanes (Spring and Summer 2012)

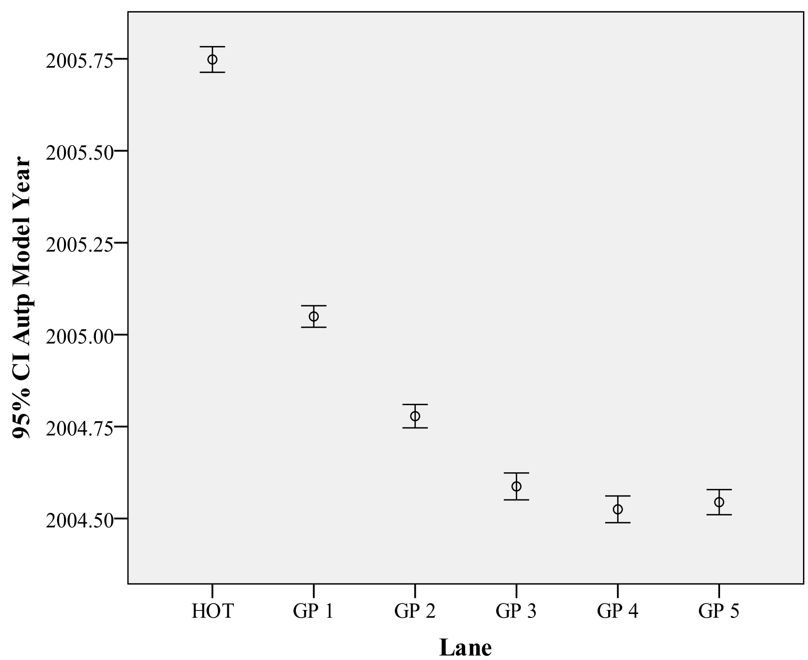


Figure 17 Confidence Interval for Vehicle Model Year across the Lanes (Spring and Summer 2012)

On average, the value of vehicles using the HOT lane is approximately $2,100 (23%) higher than the value of vehicles using the GP lanes. Similarly, the average vehicle model year is about one year higher in the HOT lane than in the GP lanes. The Mann-Whitney test rejects the hypothesis of same distribution for both HOT and GP vehicle values (sig.<0.000). Not only is the HOT lane vehicle value distribution significantly different from the general purpose lanes, the two general purpose lanes immediately adjacent to the HOT lane exhibit significantly different distributions from each other (Kruskal-Wallis test with sig.<0.000).

Table 5 (a) and (b) Vehicle Value and Model Year Statistics across Lanes

Statistics	HOT	GP1	GP2	GP3	GP4	GP5
Average	$11,382	$9,752	$9,267	$8,995	$8,971	$9,018
Median	$10,080	$8,531	$8,041	$7,692	$7,581	$7,659
St. Dev.	$7,270	$6,352	$6,173	$6,174	$6,246	$6,311

Variables	Statistics	HOT	GP	Diff	Diff (%)
Vehicle value estimate($)	Average	$11,382	$9,242	$2,140	23%
	Median	$10,080	$7,986	$2,094	26%
	St. Dev.	$7,270	$6,265	$1,005	NA*
Model Year	Average	2005.8	2004.7	1.0	NA*
	Median	2006	2005	1	NA*
	St. Dev.	3.78	4.09	0.30	NA*
Sample Size	-	45,295	288,450	NA*	NA*

* Not Applicable

Because the distributions are not normal and do not have exactly the same shape, independent sample nonparametric median test were performed instead of t-tests to compare central tendency of HOT versus GP lanes vehicle values. The Wilcoxon Mann-

Whitney test rejects the hypothesis of same median for HOT and GP vehicle value (sig.<0.000). Because no similar study was identified in the literature, no evaluation metric was available to assess the noted difference between HOT and GP lane vehicle value in terms of socioeconomic impact.

Figure 18 show the same confidence intervals illustrated in Figure 16 across different lanes, times and sites. Figure 18(b) aggregates all the GP lanes while Figure 18(a) shows each general purpose lane separately. Five data collection sites across the corridor in the order of exiting from central district are: CTR (Chamblee Tucker Road), JCB (Jimmy Carter Boulevard), BRR (Beaver Ruin Road), PHR (Pleasant Hill Rd) and OPR (Old Peachtree Road) and are previously illustrated in Figure 4.

The average vehicle value in GP lane increases toward the wealthier suburban areas (from left to right), matching the distribution of the commutershed with respect to household annual income (Figure 15). However, the HOT lane does not follow the same increasing vehicle value pattern. Instead, vehicle values across different sites are almost equal along HOT lane observation sites, except for CTR and JCB which are significantly lower in the afternoon. This may be the result of the influence of non-commuting trips such as shopping which are mainly in the afternoon around these two sites (Khoeini et al., 2012).

Comparing values across GP lanes, the same decreasing trend in vehicle value from left lane to right lane can be observed at all data collection sites, except for OPR and CTR. The reason may be that the I-285 interchange is located between CTR and JCB and Highway 316 is located between PHR and OPR which changes the vehicle lane distribution. Moreover, the OPR site is almost located at the end of Atlanta metro area and most commuters exit the corridor before reaching that point.

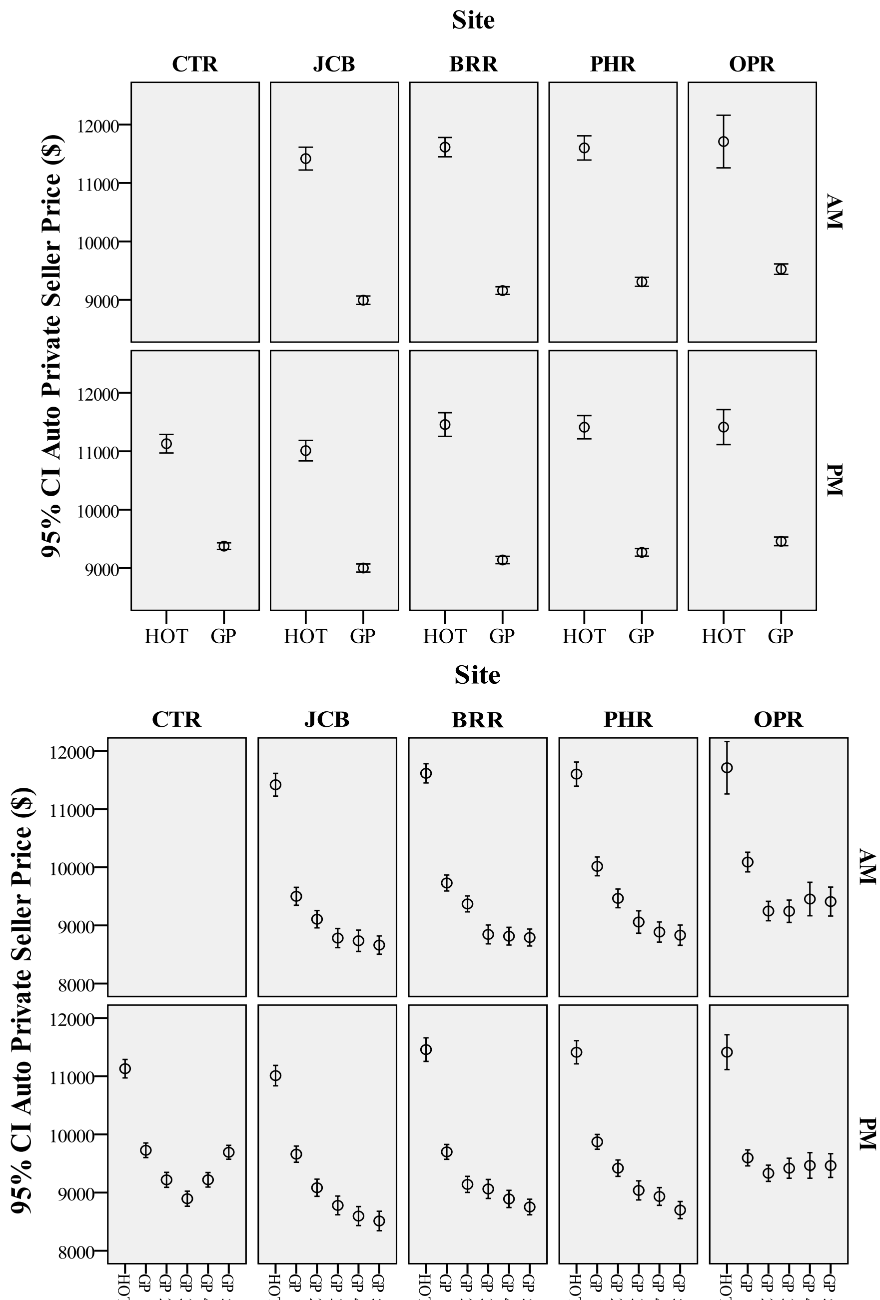


Figure 18 (a) and (b) 95% Confidence Interval for Auto Private Vehicle Seller Price across Lanes, Time and Sites (FIGURE 6(a) aggregates all GP Lanes)

5.2.3. Vehicle Rank Analysis

As discussed earlier, the average vehicle value in HOT lane is $2,140 (23%) higher than the GP lanes. Higher vehicle values may arise from two factors: 1) higher value make/model vehicles may be used in the HOT lane (e.g. more sport utility vehicles), and 2) newer vehicles within the same make/model may be used in the HOT lanes (perhaps correlated with income and ownership of more recently purchased vehicles). To assess how each factor contributes to change in average vehicle value, vehicles in GP as well as HOT lane are aggregated by make and model pair.

More than 400 unique make/model contributions (composed of a variety of model years) were observed. However, in performing statistical comparisons, only vehicle makes and models observed more than 30 times were included in the assessment, reducing the total number of unique make/model combinations to a manageable 180. Table 6 illustrates the top 20 make/model combinations by observation frequency in the HOT lane and GP lanes, along with the average model year and average vehicle value for the make/model combination.

The top four vehicles, Honda-Accord, Honda-Civic, Toyota-Camry and Ford-F150 retain their ranking in HOT and GP lanes. However, average model year is higher in the HOT lane (newer vehicles) leading to a higher vehicle value for the vehicles in the HOT lane compared to GP lanes. While Lexus vehicles do appear more frequently in the HOT lanes than in the general purpose lanes, the popular term "Lexus Lanes" is certainly a misnomer for HOT lanes. HOT lanes are Accord, Civic, Camry, and F150 lanes… just like the GP lanes.

The change in rankings for some vehicles such as Nissan-Altima or Honda-CRV is only one or two places, while some other vehicles such as the Ford-E350 or Lincoln-MKX moved more than 100 places in the ranking. The highest change in vehicle ranking belongs to Ford-E350 which is in 377th place in GP lanes and 131st place in HOT lane. While the average vehicle value didn't change for this vehicle, the main reason for the big ranking change is the body style of the vehicle which has enough space for 12 passengers and is often being used for vanpooling.

With such a large sample (and 180 unique make/model combinations), the observed change in rank order is likely to be significant. The significance of rank order change can be tested using the Wilcoxon signed-rank test, which is a non-parametric test used when comparing two related samples, using the ranks of the pairs of scores formed by the matched pairs in the sample (Daniel, 1990). The Wilcoxon signed-rank test rejects the null hypothesis that the median of differences between different vehicles rate of presence in GP versus HOT lanes equals zero at the 0.046 significance level. After becoming statistically confident that there is a significant change in vehicle rank orders, the next step is to assess how much of the difference in vehicle value at HOT versus GP lanes is caused by the vehicles rank order change compared to differences in model year for the same make-model combinations.

As stated earlier, there is $2,140 difference between HOT and GP vehicle value, which is 23% higher than the average GP vehicle value. Multiplying the difference in observed percentages in HOT versus GP lanes by GP vehicle values (for each vehicle make and model combination) and summing the values, equals $882. This amounts to 42% of the $2,140 difference, which means vehicles rank order changes account for 42%

of the change in vehicle value. The remaining 58% results from the one year increase in model year (on average) in HOT lane compared to general purpose lanes. To summarize, of 23% difference in vehicle value between HOT and GP lanes, 13% results from an increase in average model year and 10% is results from the change in vehicle makes/models (rank orders).

Table 6 Rank Order of Vehicle Make and Model Usage in HOT and GP Lanes

Rank	HOT				GP			
	Vehicle Make	Vehicle Model	Average Model Year	Average Vehicle Value($)	Vehicle Make	Vehicle Model	Average Model Year	Average Vehicle Value($)
1	HOND	ACCORD	2004.6	9,585	HOND	ACCORD	2003.5	8,189
2	HOND	CIVIC	2005.0	8,130	HOND	CIVIC	2004.3	7,270
3	TOYT	CAMRY	2005.2	8,840	TOYT	CAMRY	2004.3	7,878
4	FORD	F150	2005.8	9,353	FORD	F150	2004.3	7,486
5	NISS	ALTIMA	2006.9	9,001	TOYT	COROLLA	2005.1	8,031
6	HOND	CR-V	2005.8	10,529	NISS	ALTIMA	2005.9	8,103
7	TOYT	COROLLA	2006.1	7,942	CHEV	SILVERADO	2005.0	8,295
8	INFI	G35	2005.0	11,236	HOND	CR-V	2005.2	9,765
9	FORD	ECONOLIN E	2007.4	9,560	FORD	EXPLORER	2002.6	5,805
10	CHEV	SILVERAD O	2005.3	8,421	FORD	ECONOLINE	2004.5	7,031
11	FORD	EXPLORER	2004.0	8,078	DODG	RAM	2004.9	7,734
12	HOND	ODYSSEY	2006.4	12,013	FORD	MUSTANG	2004.3	8,890
13	LEXS	RX	2004.6	15,501	HOND	ODYSSEY	2004.7	9,425
14	FORD	MUSTANG	2005.4	10,700	NISS	MAXIMA	2002.5	8,072
15	JEEP	CHEROKEE	2005.0	9,894	TOYT	TACOMA	2004.6	11,076
16	NISS	MAXIMA	2005.0	12,088	JEEP	CHEROKEE	2002.8	6,833
17	LEXS	ES	2004.6	13,425	TOYT	4RUNNER	2002.8	9,236
18	CHEV	TAHOE	2006.1	15,650	CHEV	EXPRESS	2005.5	8,778
19	HOND	PILOT	2006.7	12,252	TOYT	SIENNA	2004.7	8,941
20	TOYT	4RUNNER	2004.8	12,530	INFI	G35	2005.0	11,373

5.2.4. Users Income and Vehicle Value Relationship

While the purpose of this section is to assess the potential use of vehicle value as a proxy for income, this section will examine the users' household annual income to assess the reliability of the vehicle value methodology presented in this section. Household level income data were procured from self-reported marketing data (4.1.4).

The first analysis in this section is the correlation between average household annual income and vehicle value for 2,700 unique vehicles make/model/year combination in the study. For each vehicle type, average annual income of all the households that have that specific vehicle type was assigned. The Spearman correlation coefficient, which is a nonparametric alternative to Pearson correlation coefficient, between vehicle value and household annual income is 0.422 (sig. < 0.0001). At first glance, the correlation is positive and significant, but may not be high enough to justify using vehicle value in place of household annual income. However, when self-reported household annual income data were employed in the same analysis conducted in this study, almost the exact same trend in differences across the lanes was observed as was observed in the previous section employing vehicle value.

Figure 19 shows mean confidence interval for household annual income across the lanes. Comparing Figure 19 to Figure 16, nearly the same decreasing trend in vehicle value and income from left lane to right lane can be observed.

Table 7 (a) and (b) also show the statistics for household annual income across the lanes. Specifically, Table 7 (a) shows average, median and standard deviation of household annual income across all the lanes. Similar to the mean, the median value also decreased from left lane (HOT) to right lane. Standard deviation is consistent across the lanes, with slightly higher deviations in the HOT lane. Table 7 (b) has calculated the difference and percent difference of household annual income between HOT and aggregated general purpose lanes.

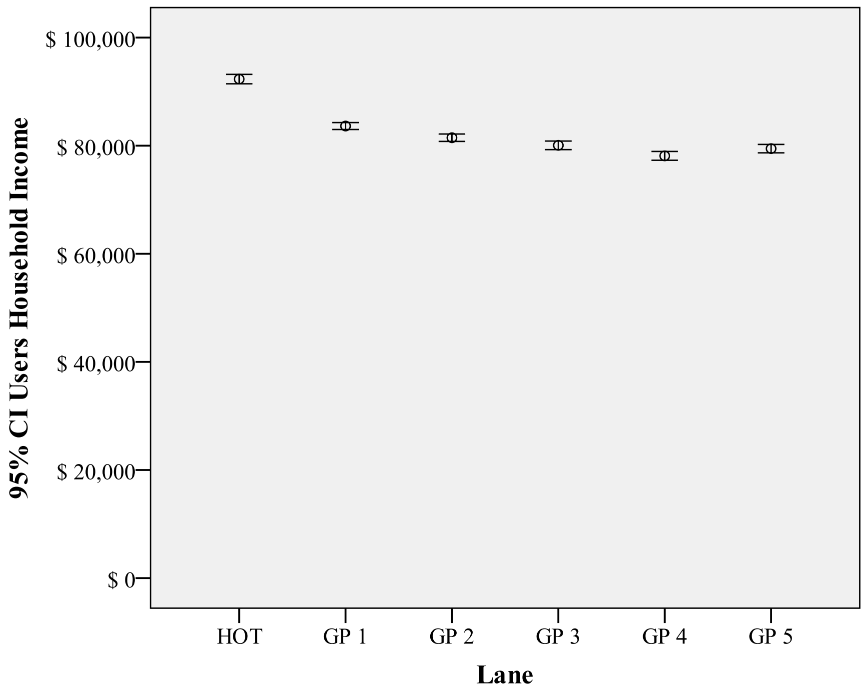


Figure 19 95% Confidence Interval for Users' Household Annual Income across Lanes

While the average household annual income along HOT lane is about 14% higher compared to the general purpose lanes, the HOT lane average vehicle value is about 23% higher. Similarly, the difference in median household annual income is about 11% while the same median difference in vehicle value is 26%. Vehicle value shows the same statistically-significant increasing and decreasing trends noted for household annual income. Certainly, the relative percent differences in vehicle value should not be used to directly substitute for percent differences in household annual income, but further studies should be able to reveal a proper surrogate relationship.

To better demonstrate the agreement and association between vehicle value and household annual income, the last analysis of this section compares the distribution across the lanes. For comparative illustration, quartile distributions of vehicle value and household annual income have been presented across the lanes in Figure 20 and Figure

21. Quartile boundaries have been calculated using the entire data regardless of the lane assignment (last column).

Table 7 (a) and (b) Users Household Annual Income Statistics across Lanes

Statistics	HOT	GP1	GP2	GP3	GP4	GP5
Average	$ 92,320	$ 83,637	$ 81,475	$ 80063	$ 78,119	$ 79,455
Median	$ 87,500	$ 72,500	$ 72,500	$ 72,500	$ 67,500	$ 67,500
St. Dev.	$ 47,984	$ 44,201	$ 44,188	$ 44,072	$ 44,724	$ 45,424

Variables	Statistics	HOT	GP	Diff	Diff (%)
Users Household Annual Income ($)	Average	$ 92,320	$ 80,875	$ 11,445	14.15%
	Median	$ 87,500	$ 72,500	$ 15,000	11.43%
	St. Dev.	$ 47,984	$ 44,532	$ 3,452	NA*
	Sample Size	11,618	70,910	NA*	NA*

* Not Applicable

Vehicle value quartile boundaries are: lower than $4,703 (vehicle value low), between $4,703 and $8,279 (vehicle value medium), between $8,279 and $13,100 (vehicle value high), and higher than $13,100 (vehicle value very high). The probability of a vehicle belongs to any of the four categories, without any information about the lane assignment, is 0.25. Household annual income quartile boundaries are: lower than $57,500 (income low), between $57,500 and $92,500 (income medium), between $92,500 and $127,500 (income high), and more than $127,500 (income very high). Similarly, the probability of a household belongs to any of the four categories, without any information about the lane assignment, is 0.25.

As a result, the high level of agreement between vehicle value and household annual income can be seen by visually comparing the two graphs. Non-parametric statistical test to compare the distributions have been employed. Household annual income distributions across all the lanes are significantly different except the most two right lanes (Mann-Whitney U Test with sig. < 0.000). Vehicle value distributions across all the lanes are significantly different except the most three right lanes (Kruskal-Wallis Test with sig. < 0.000).

Regarding income distribution along the HOT lane, 34% of the users belong to the top quartile (9% more than average) and 27% belongs to the 3rd quartile (2% more than average); whereas, 22% belong to the 2nd quartile (3% less than the average) and 17% belong to the lowest quartile (8% less than average). In other words, 11% of the users out of the two lowest income quartiles chose to use the general purpose lanes and were replaced with households from the two upper quartiles after conversion to a HOT lane.

Regarding the vehicle value distribution along the HOT lane, 32% percent of the vehicles belong to the top quartile (8% more than average) and 27% belongs to the 3rd quartile (1% more than average); whereas, 20% belong to the 2nd quartile (3% less than average) and 21% belong to the lowest income quartile (6% less than average). That is to say, 9% of the of the vehicles out of the two lowest vehicle value quartiles chose not to use the HOT lane and replaced with vehicles from the two upper quartiles after conversion to a HOT lane. The results using household income and vehicle value are in good agreement.

The statistical test results indicate that, the three right GP lanes have very similar distribution with 3% more than average belong to low income and low vehicle value. The GP1 and GP2 lanes, which are directly adjacent to the HOT lane, are closer to the general population quartile distribution than the other GP lanes and the HOT lane.

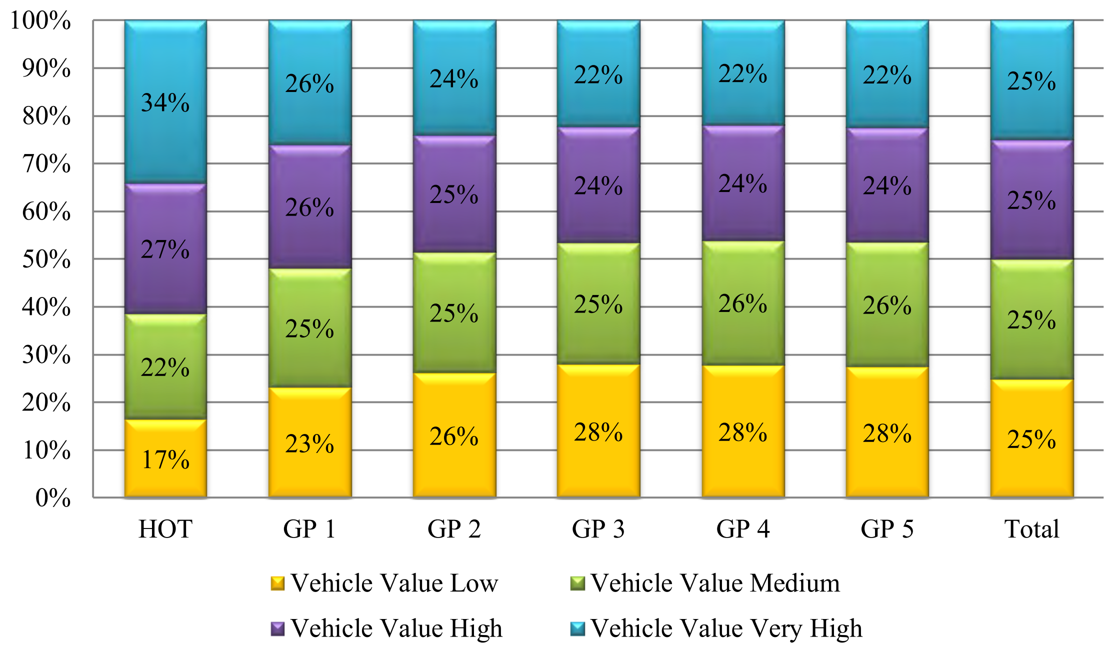


Figure 20 Vehicle Value Quartile Distribution across the Lanes

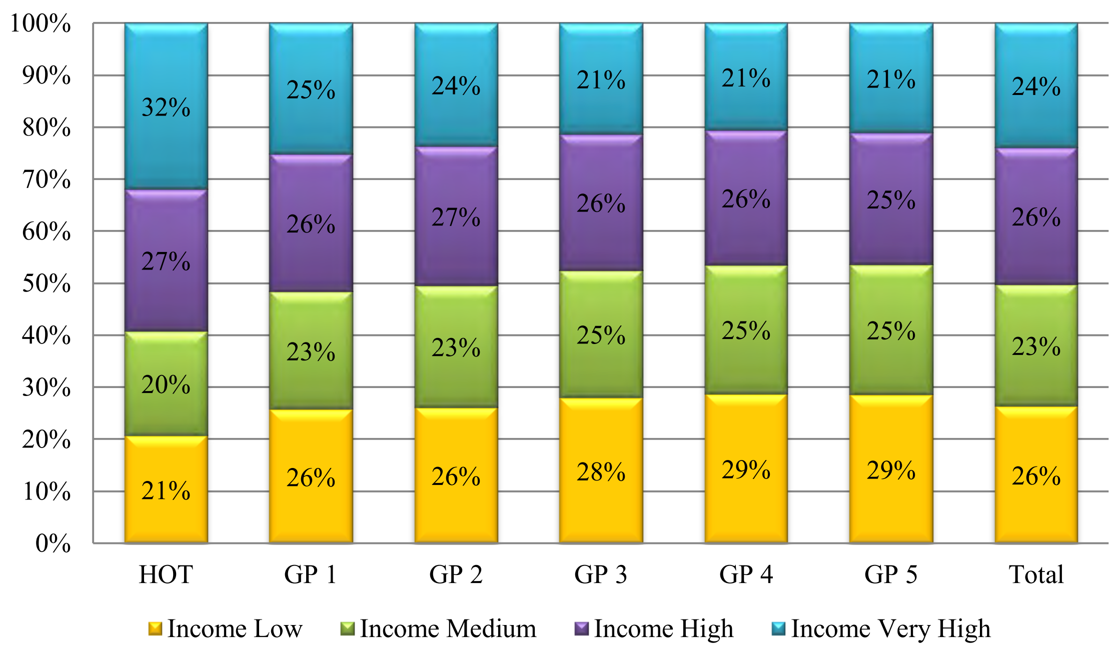


Figure 21 Household Annual Income Quartile Distribution across the Lanes

Lastly, this study once again rejects the concept of "Lexus Lane" by illustrating that significant number of low income users are using the HOT lane; however, very high income travelers (more than $127,000 household annual income) are using HOT lane twice as frequent as the low income travelers (less than $57,500 household annual income). Finally, considering the low cost of the proposed method, vehicle value analysis via license plate data collection demonstrates the potential to be applied in practice in similar studies.

5.3. Conversion Impact on HOV Carpoolers[3]

The main user group negatively impacted by the HOV-to-HOT conversion was HOV two-person carpoolers, who used to be able to use the carpool lanes for free and now have to either split a toll or find a third person to join their carpool if they want to continue using the HOT lane. This section examines data collected pre- and post-conversion from a HOV to HOT facility to assess changes in the demographic characteristics of these user groups.

The HOT trip summary data, collected along HOT Lane (applicable for only PeachPass holders after the conversion), is utilized to investigate the travel frequency and willingness to pay of carpool market. Finally, the response of the observed carpool market to the pricing is investigated as a function of their demographic attributes.

The HOV lane license plates which appeared frequently over the one year study period (before conversion) were identified. This identification was based on two

---

[3] This section has been published: S. Khoeini, A. Sheikh. "The Fate of HOV Users after Pricing: Atlanta's I-85 HOV2 to HOT3 Conversion"; *2013 Freeway & Managed Lane Operations Meeting and Conference*, Atlanta, GA, June 2013 (Best section award).

variables: 1) the number of times that the plate was observed in the HOV lane, and 2) the percentage of observations the vehicle was observed in the HOV lane (i.e. HOV lane use vs. general purpose lane use). More than 3,000 HOV plates were observed in the HOV lane five or more times over one-year period and used the HOV lane more than 80% of the time. These license plates were then matched to the following data streams for analysis.

Georgia Tech receives a daily file of HOT trips from the State Road and Tollway Authority (SRTA) and its contractor, ETC Corporation. Each file contains records of the previous day's Express Lane trips. These records provide data such as trip entry and exit time, start and end points, amount of toll paid, and the mode ("Toll" or "Non-Toll") that the vehicle transponder was operating in. This study used trip records from the start of operations in October 2011 until the end of the 2012 calendar year (15 months). Of the plates that were identified as frequent HOV users, 1,724 returned valid results from this data source for a total of 277,000 trips.

5.3.1. HOV User Socio-economic Characteristics

The distributions of corridor users' socio-economic characteristics such as household size, income, vehicle availability, ethnicity, age, and education have been retrieved based on household level marketing data (section 4.1.4) for more than 273,307 households. From these households, 3,000 households have been identified as frequent HOV users, as discussed earlier, and their socio-economic distributions are presented mutually, in the same graphs, to better visualize the differences in Figure 22. The vertical axes of the graphs are all expressed as percentages becuae the sample sizes of two groups are different.

In terms of income, the HOV frequent users' income distribution shows slightly higher incomes for HOV users compared to the rest of the corridor users. Starting from households with $62,500 income, the percentages of HOV users are higher while these percentages are lower up to this value. This difference in distributions corresponds to a $6,700 (8% of average income) difference in average income.

In terms of vehicle availability, HOV frequent users have a lower percentage of 1, 4, and 5 vehicles per household, and a higher percentage of 2 and 3 vehicles per household. This corresponds to a higher average vehicle availability value (the difference is 1.14) for frequent HOV users compared to general corridor users. Household size was slightly larger for frequent HOV users, with 2 or more members and fewer single person households in the frequent HOV users. This was expected because of the higher probability of carpooling with household members (fampooling).

Regarding ethnicity, Hispanic and Asian households make up a significantly higher percentage of HOV users, while White and African-American households make up a lower percentage of HOV users. In terms of education, the frequent HOV users are slightly more educated with respect to some college and college degrees, but fewer graduate degrees. Lastly, regarding the age of the head of the household, HOV users have more frequency in the 35 to 54 year-old range and lower frequency for higher and lower ages compared to general corridor users.

In summary, the frequent HOV users who may have been negatively impacted by the conversion were generally households represented by middle-aged head-of-household demographics with slightly higher household size, income, vehicle availability, and education level compared to the general corridor users.

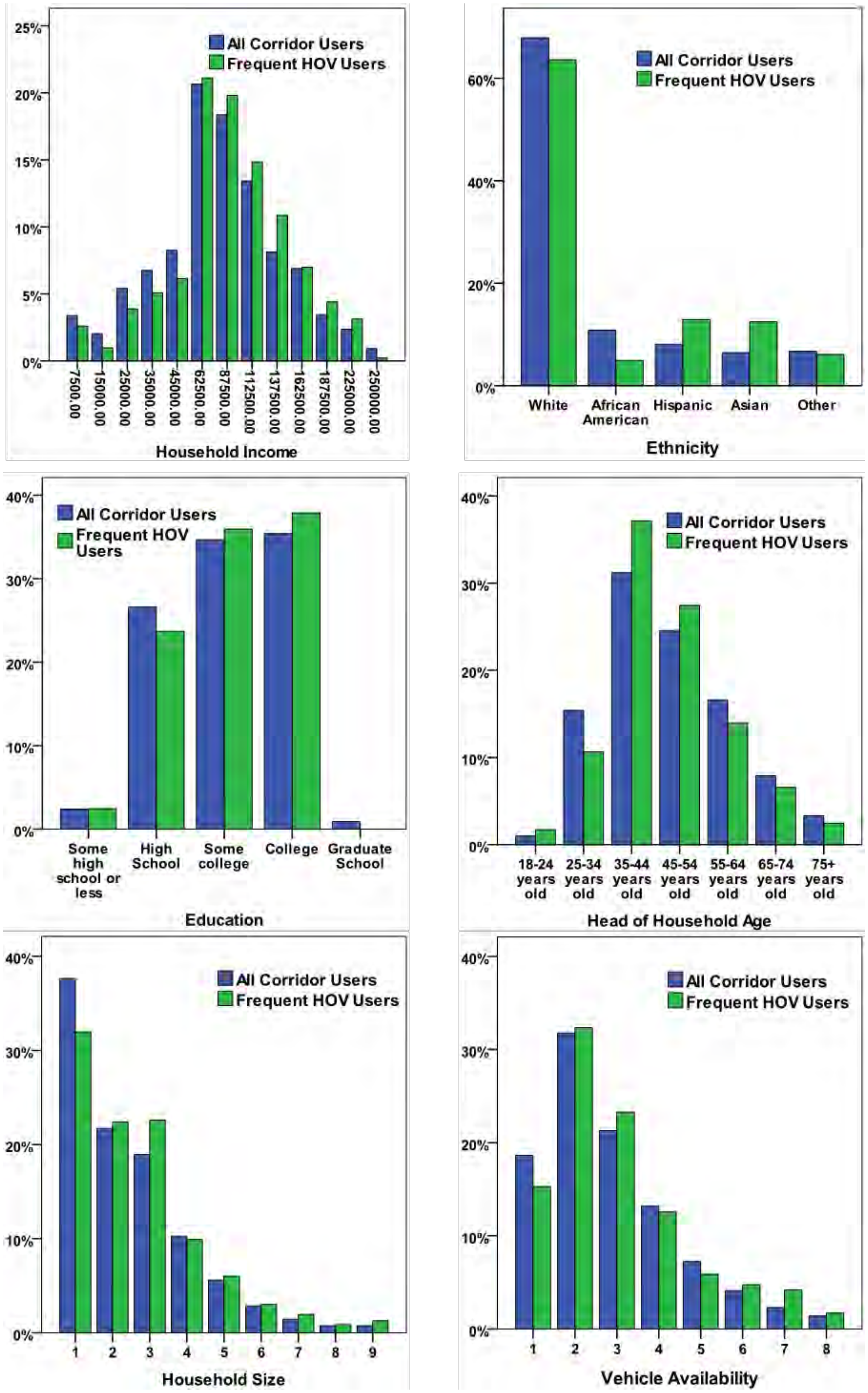


Figure 22 HOV Frequent Users (n=3,050) vs. General Corridor Users (n=270,257) Characteristics

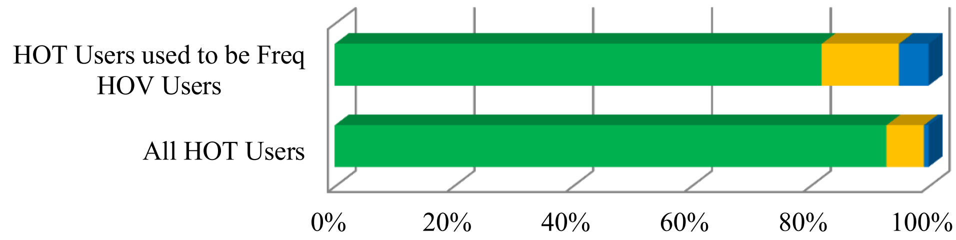
5.3.2. HOV Users Response to Pricing

5.3.2.1. Lane Assignment

In this section, the frequent HOV users' response to pricing will be evaluated using trip summary data. Of the 3,055 HOV frequent users license plates, only 1,726 plates (56%) were matched to at least one trip in the HOT lane after the conversion. This implies than the remaining 44% are either only using general purpose lanes or they have changed their commute route or mode, with some vehicles also having been scrapped or sold outside of the region.

The frequent HOV users, who were observed in the HOT lane, exhibit higher frequency of HOT lane use, compared to all HOT lane users. The median number of HOT trips per plate for all users is 2 trips/month (mean equals 2.8), while the frequent HOV users that still use the HOT lane are making a median of 5 trips/month (mean equals 10.7).

In examining the most frequent HOT users (top 20%), those who used to be frequent HOV users (345 plates that make at least 18 HOT trips per month), 82% are making only toll-mode trips (with either one-person or two-persons in the vehicle), 13% are making non-tolled trips (3+ vehicle occupancy) and the remaining 5% are switching between toll and non-tolled trips. Unfortunately, it is not possible from the data collected to determine whether these 2-person carpools broke apart into SOVs or stayed together and are splitting the toll. The same percentages were calculated for all HOT users and the results are shown in Figure 23. Frequent HOV users make non-toll trips (3+ persons) with more than twice the frequency of general HOT users, and they switch between the toll and non-toll modes with five times the frequency of general HOT users.



	All HOT Users	HOT Users used to be Freq HOV Users
HOT Toll Mode	93%	82%
HOT Non-Toll Mode	6%	13%
HOT Both Toll and Non-Toll	1%	5%

Figure 23 HOT Lane Toll Mode Distributions

5.3.2.2. Willingness to pay toll

In terms of willingness to pay toll, the average toll amount per toll trip for all HOT users is $1.25 while for frequent HOV users is $1.43. This implies that frequent HOV users are willing to pay 13% more per toll trip on average. Interestingly, the estimated trip length of frequent carpoolers is longer than the general HOT user. Hence, frequent HOV users are paying 16.7 cents per mile on average for toll trips and their average HOT lane trip length is 8.45 miles. On the other hand, general HOT users are paying 15.9 cents per mile on average of toll trip and their average toll trip length is 7.4 miles. While the average toll paid by frequent HOV users is 13% higher than that of all HOT users, their average per-mile toll amount is only 5% higher. However, the fact that 44% of frequent HOV users are not using the HOT lane any more should have also been considered in future Traffic and Revenue Studies.

5.3.2.3. Socioeconomic analysis

After assessing the frequent HOV users' response to pricing, this section investigates the relation between their response to pricing and their socio-economic attributes. Based on previous analysis, frequent HOV users fall in five categories in terms of response to pricing:

1. Frequent HOT users who are making only non-toll trips;
2. Frequent HOT users who are making both toll and non-tolled trips;
3. Frequent HOT users who are making only tolled trips;
4. Infrequent HOT users (fewer than 10 times per month);
5. Vehicles no longer observed in the HOT lane (users that may have switched to general purpose lanes, other modes, other routes, or obtained new vehicles).

The socio-economic attributes across these groups for HOV frequent users are presented in Figure 24. Group 2 (frequent HOT users that are making both toll and non-tolled trips) has an insufficient sample size (lower than 30 plates) and was eliminated. Moreover, since the data are not normally distributed, box plots have been used for analysis, with the median as the center value.

In terms of vehicle availability, all HOT users are very similar while the last group (which stopped using the managed lane after conversion) has more households with high vehicle availability. This may indicate the presence of households who found no incentive to continue to carpool while they have enough vehicles to drive alone. In terms of household size, as was expected non-toll trip makers have a higher median household size compared to other groups and the 'Other Lane/Mode/Route' group has fewer households with a household size greater than 2.0.

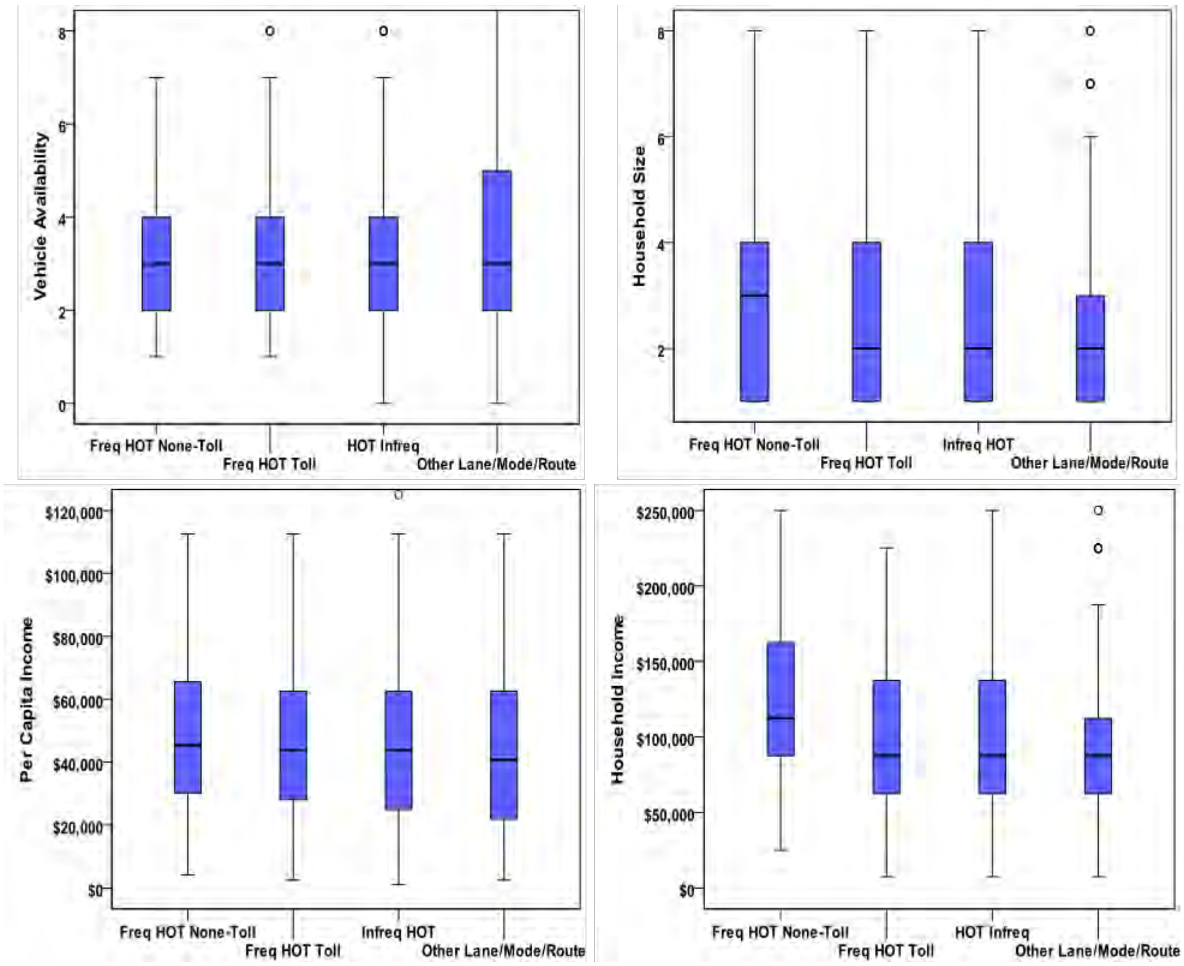


Figure 24 Socio-economic Attributes across Groups of Frequent HOV Users

The household income graph shows different income distributions across the groups, with higher income for non-toll mode trip makers. However, the per capita income (total income divided by household size) graph shows a more similar distribution across the groups. This difference may also tie back to the positive correlation of 0.3 between income and household size.

Age and education attributes do not appear to differ significantly across these user groups. However, the impact of ethnicity is very interesting and has been presented in Figure 25. If response across each ethnic group were the same, the bars would be expected to be level within each ethnic cluster. White former HOV users are more likely to use the lane in either toll or non-toll modes than to change lanes, modes, or routes. African American and Hispanic users illustrate the opposite trend and their vehicles show up less frequently in the HOT lanes (albeit, the fate of vehicles over a one-year period has yet to be investigated and may differ across ethnic groups with respect to vehicle turnover). Lastly, Asian users may have already had a higher percentage of 3-person carpools in the HOV lanes, or may have been more willing to form 3-person carpools (increase their occupancy) after HOT conversion.

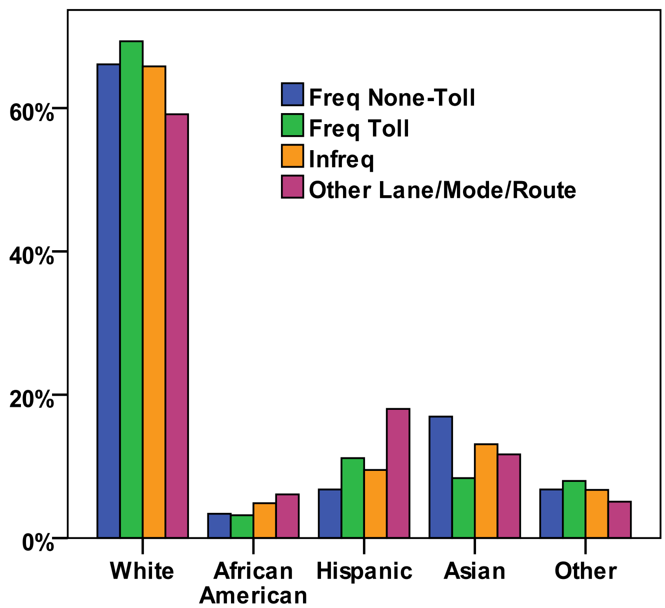


Figure 25 Response to HOT Lane Implementation among Different Ethnic Groups of HOV Frequent Users

In summary, no negative impact was observed by income or vehicle availability for frequent HOV users when it came to their travel choice after conversion from HOV2+ to HOT3+. However, larger household sizes may be more willing to make non-tolled trips (3-person occupancy) on the HOT lane.

5.4. Block group Level Analysis[4]

The section assesses and analyzes changes in block groups' HOV, HOT lanes frequency of use, in relation to their spatial location with respect to the corridor and socioeconomic attributes.

Because survey data are expensive and not feasible for many tranportaton projects, this section uses publicly available American Community Survey data at the block group level according to the well-defined commutershed. While publicly available data such as American Community Survey do not provide household-level accuracy, they are costless and accurate at certain geographic boundaries. Whenever statistically significant spatial information about corridor users are avaialable, these data sources have the potential for travel behaiviour and socioeconomic studies. The main purpose of this section is to understand the relationship between the block groups socio-spatial characteristics and managed lane usage behavior.

5.4.1. Exploring Block group Level Managed Lane Usage

The aggregated travel related attributes for the 2,102 block groups that are intersected with the I-85 corridor commutershed (95% directional distributional ellipse), have been acquired for analysis. Table 8 illustrates the average frequency of observation along the corridor, HOV lane, and HOT lane per block group. The average HOV and HOT lanes use frequencies normalized by number of block groups' workers as well as

---

[4] This section has been published: S. Khoeini, R. Guensler. "Socio-spatial Model for Managed Lane System Expansion based on a Before and After Study of Atlanta I-85 HOT Lane"; 60th Annual North American Meetings of the Regional Science Association International, Atlanta, GA, Nov 2013.

corridor frequencies are also illustrated per block group. "HOV Usage" refers to percent of trips per block group, were observed along HOV lane; and, similarly "HOT Usage" refers to percent of trips per block group, were observed along HOT lane.

Table 8 Blockgroup Level Travel Related Statistics (N=2,102 Block groups)

	Average	Std. Deviation
Observation Freq per BG	5097.15	4,838.60
HOV Observation Freq per BG	365.98	360.16
HOT Observation Freq per BG	287.32	312.56
HOV Observation Freq per BG Workers	0.21	0.07
HOT Observation Freq per BG Workers	0.16	0.14
HOV Usage	0.11	0.05
HOT Usage	0.13	0.06

The average HOV observation frequency is larger than the average HOT observation frequency because of the larger amount of data collection before the conversion. However, controlling for different amount of data collection before and after the conversion, average block group HOT lane usage (0.13) is larger than HOV lane usage (0.11). This implies an increase in managed lane usage per block group and coordinates well with the observed increase in managed lane throughput.

To investigate the relationship between block groups socioeconomic attributes and managed lane usage, Table 9 illustrates HOV lane and HOT lane usage median, mean and means confidence intervals across socioeconomic categories. Similarly, the following figures illustrate the managed lane usage confidence intervals across socioeconomic categories. In general, HOV lane usage is more consistent across block groups with various socioeconomic attributes compared to the HOT lane usage.

Table 9 HOV/HOT Lanes Usage across Block Groups Socioeconomic Categories (N=2,102)

		HOV Usage				HOT Usage			
		CI Low	CI High	Mean	Median	CI Low	CI High	Mean	Median
HH Size	**<2**	0.09	0.13	0.11	0.08	0.07	0.13	0.10	0.08
	2-3	0.11	0.12	0.12	0.11	0.11	0.12	0.12	0.12
	3<	0.11	0.11	0.11	0.11	0.14	0.14	0.14	0.15
Income ($1000)	**<50**	0.12	0.13	0.12	0.11	0.08	0.09	0.08	0.08
	50- 100	0.11	0.11	0.11	0.11	0.13	0.14	0.14	.15
	100<	0.10	0.11	0.11	0.10	0.15	0.17	0.16	0.18
Age (years)	**<30**	0.11	0.12	0.11	0.11	0.08	0.10	0.09	0.08
	30-40	0.11	0.11	0.11	0.11	0.13	0.14	0.14	0.15
	40<	0.11	0.14	0.12	0.11	0.10	0.12	0.11	0.10
Vehicle Ownership	**<1.5**	0.10	0.12	0.11	0.09	0.07	0.10	0.08	0.06
	1.5-2.5	0.11	0.11	0.11	0.11	0.13	0.13	0.13	0.14
	2.5<	0.11	0.12	0.12	0.11	0.14	0.15	0.14	0.16
High Education (%)	**<10**	0.11	0.12	0.12	0.11	0.11	0.12	0.12	0.13
	10- 20	0.11	0.11	0.11	0.11	0.14	0.14	0.14	0.15
	20 <	0.11	0.12	0.11	0.12	0.13	0.15	0.14	0.14
Female (%)	**<47**	0.11	0.12	0.12	0.11	0.10	0.11	0.10	0.10
	47- 53	0.11	0.12	0.11	0.11	0.13	0.14	0.14	0.15
	53<	0.11	0.12	0.11	0.11	0.12	0.13	0.12	0.12
Commute Travel Time (min)	**<30**	0.12	0.14	0.13	0.11	0.10	0.12	0.11	0.10
	30-40	0.11	0.11	0.11	0.11	0.13	0.14	0.13	0.14
	40<	0.11	0.11	0.11	0.11	0.12	0.14	0.13	.15
Drive-to-Work (%)	**<85**	0.11	0.12	0.12	0.11	0.11	0.13	0.12	0.12
	85-95	0.11	0.11	0.11	0.11	0.13	0.14	0.13	0.15
	95<	0.12	0.13	0.12	0.11	0.11	0.12	0.11	0.11
Work-at-Home (%)	**<3%**	0.11	0.12	0.12	0.11	0.09	0.10	0.10	0.10
	3-10	0.11	0.11	0.11	0.11	0.13	0.14	0.14	0.15
	10<	0.11	0.12	0.12	0.11	0.14	0.16	0.15	0.17

In terms of block group average household size, HOT lane usage increases as household size increases, whereas HOV lane usage doesn't change significantly (Figure 26). In terms of block group median household income, HOT lane usage increases with a large slope while HOV lane usage decreases slightly (Figure 27). In terms of block group average age, mid-life (30-40 years old) block groups HOT lane usage is highest;

however, no significant change in HOV lane usage across different age groups has been observed (Figure 28).

In terms of block group average vehicle ownership, HOT lane usage increases sharply and HOV lane usage doesn't change as average vehicle ownership per block group increases (Figure 30). In terms of percent of adults with graduate education per blockgroup, HOT lane usage slightly increases and HOV lane usage doesn't change as the percent of higher educated adult per block group increases (Figure 31). In terms of percent of female population, HOT lane usage is highest at the middle female ratio category (family oriented block groups) and HOV lane usage is not significantly changing across the categories (Figure 29).

In terms of average block group commute travel time, HOV lane usage is lower while HOT lane usage is higher at the 30-40 min travel time category compared to less than 30 min travel time category (Figure 32). In terms of drive to work ratio, HOV usage is highest at block groups with 95% or more drive to work ration, while HOT usage is highest at block groups with 85% to 95% drive to work ratio (Figure 33). In terms of work at home ratio, HOT lane usage increases significantly as work at home ratio increases per block group while HOV lane usage does not change significantly.

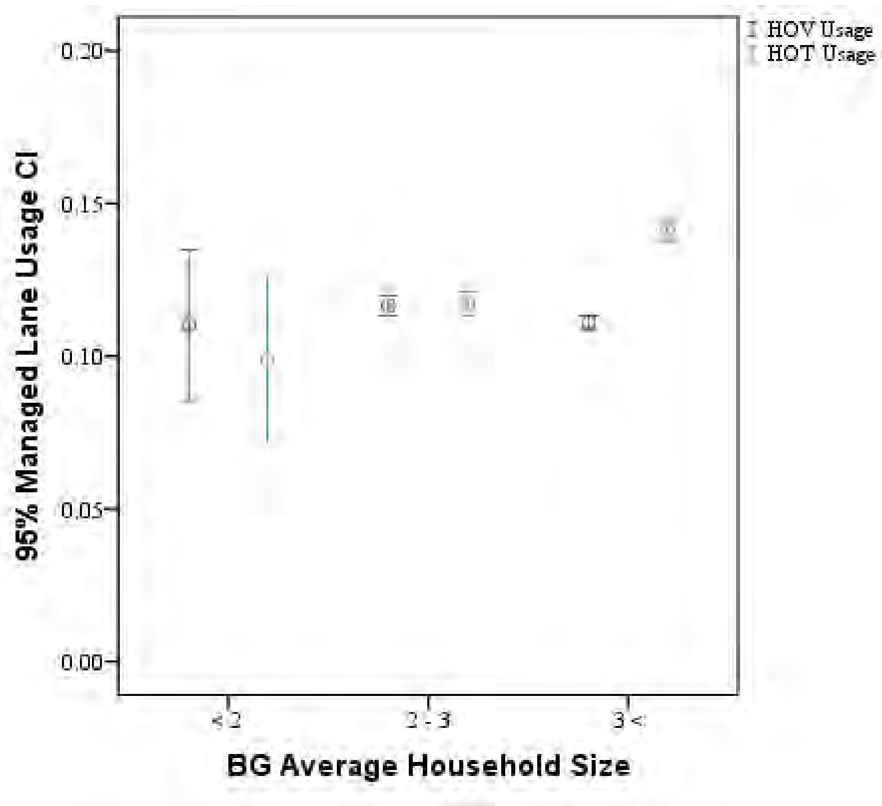


Figure 26 Block Groups Managed Lane Usage across Household Size Categories (N=2,102)

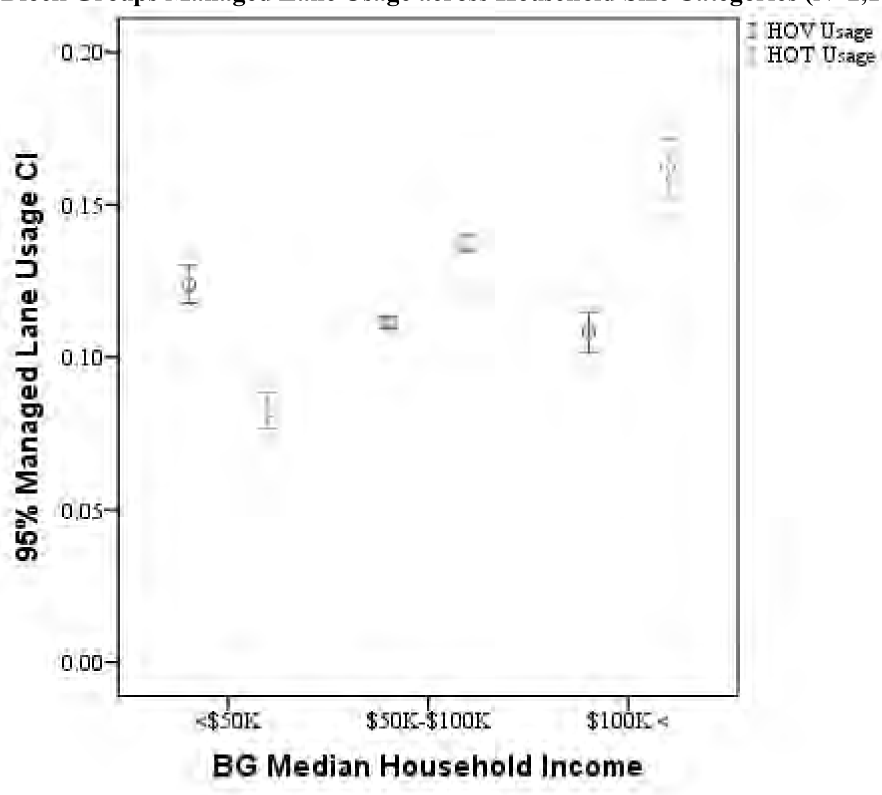


Figure 27 Block Groups Managed Lane Usage across Household Income Categories (N=2,102)

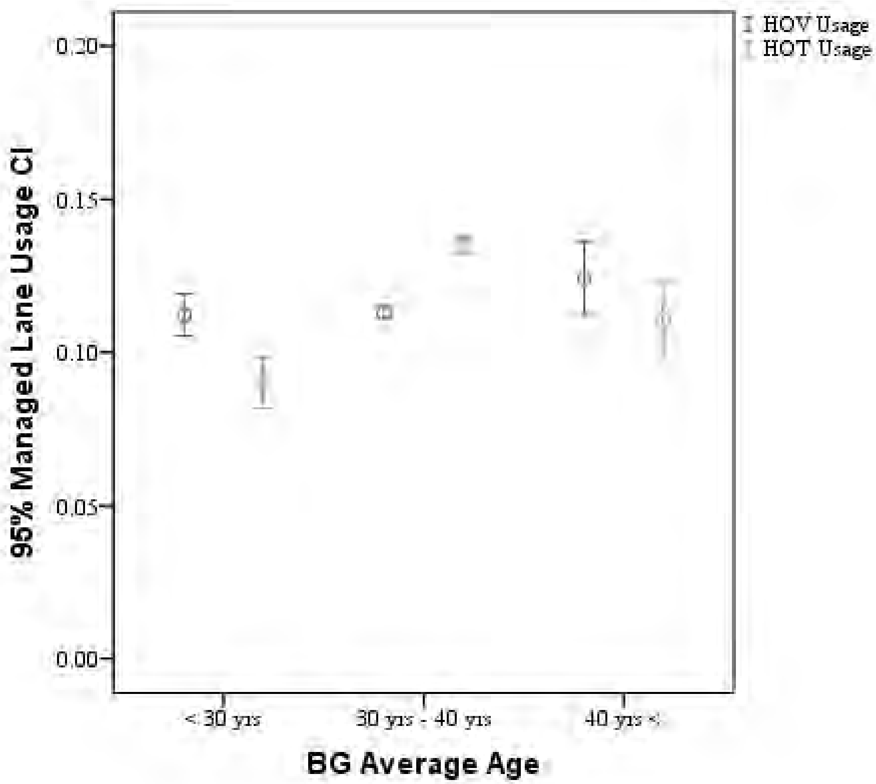


Figure 28 Block Groups Managed Lane Usage across Age Categories (N=2,102)

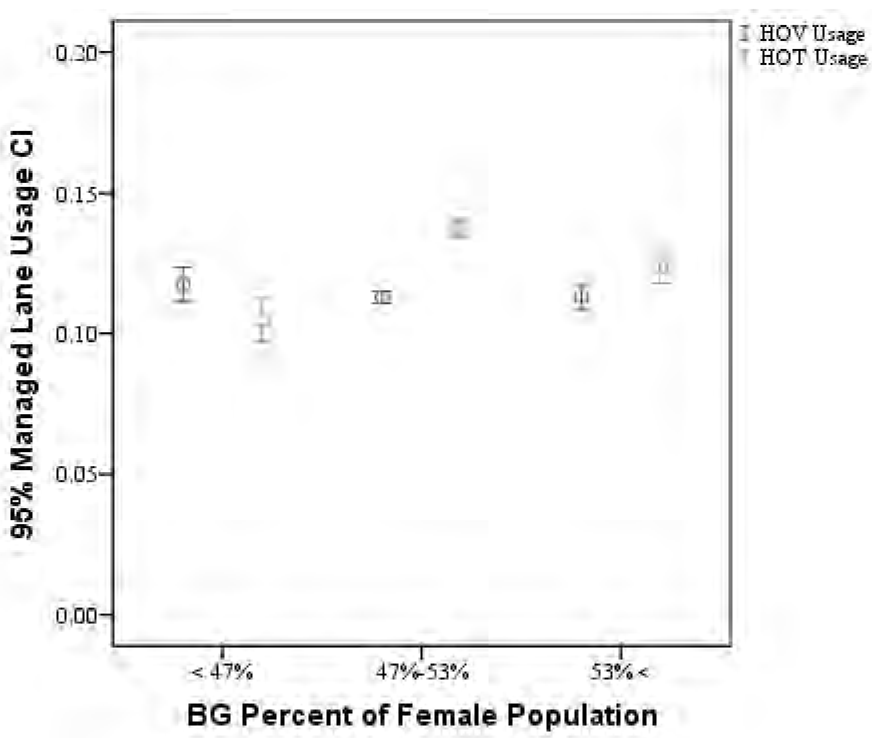


Figure 29 Block Groups Managed Lane Usage across Percent of Female Population (N=2,102)

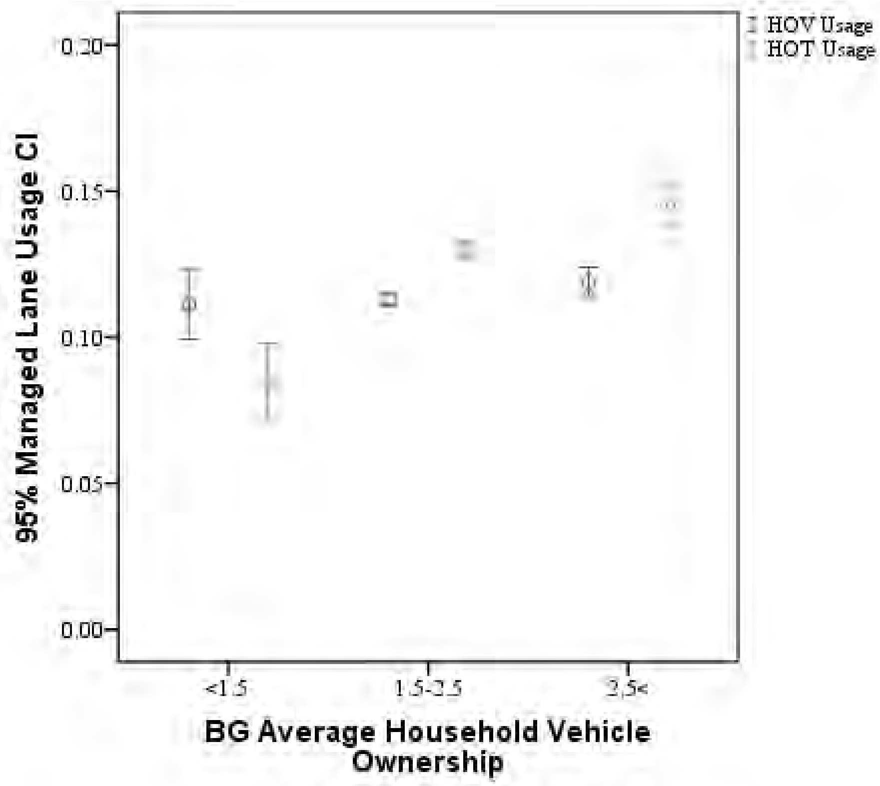


Figure 30 Block Groups Managed Lane Usage across Household Vehicle Ownership (N=2,102)

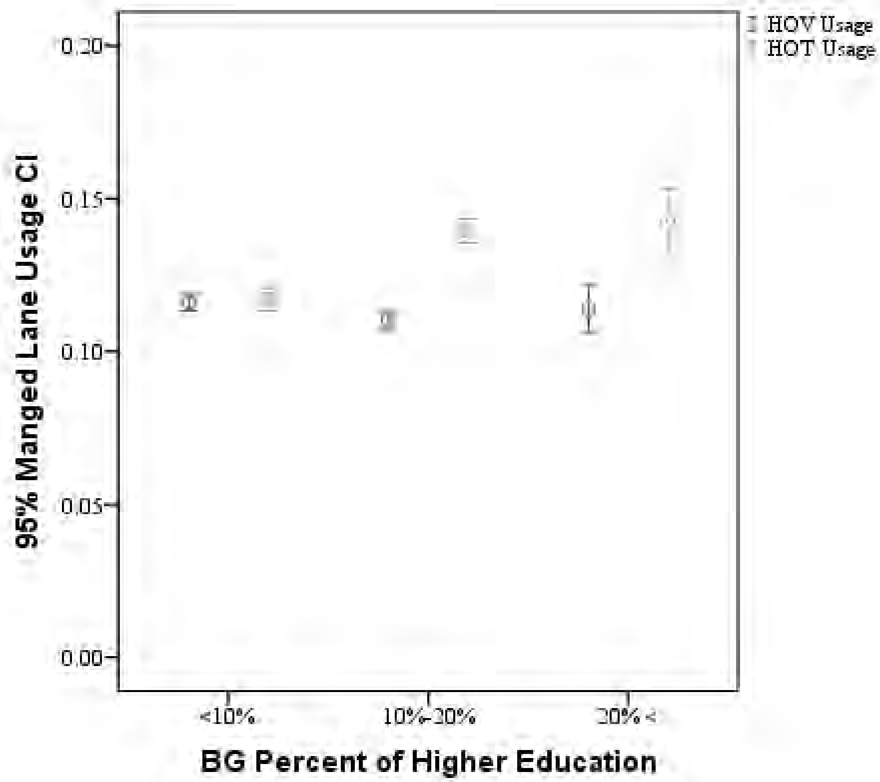


Figure 31 Block Groups Managed Lane Usage across Percent of Adults with Graduate Education Level (N=2,102)

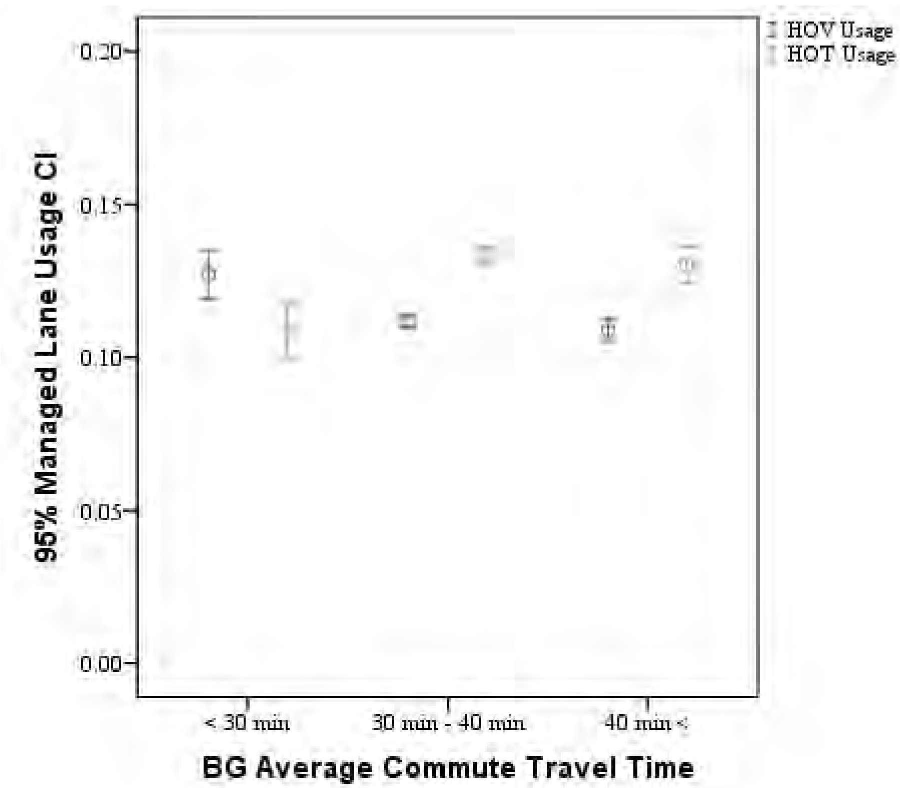


Figure 32 Block Groups Managed Lane Usage across Commute Travel Time Categories (N=2,102)

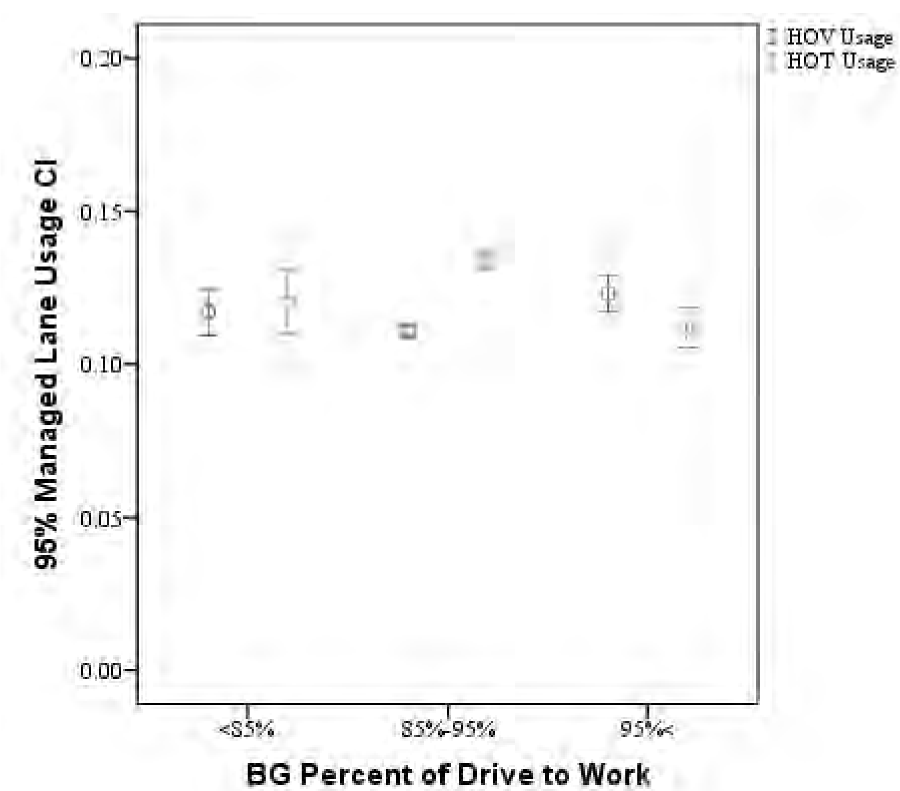


Figure 33 Block Groups Managed Lane Usage across Drive to Work Ratios Categories (N=2,102)

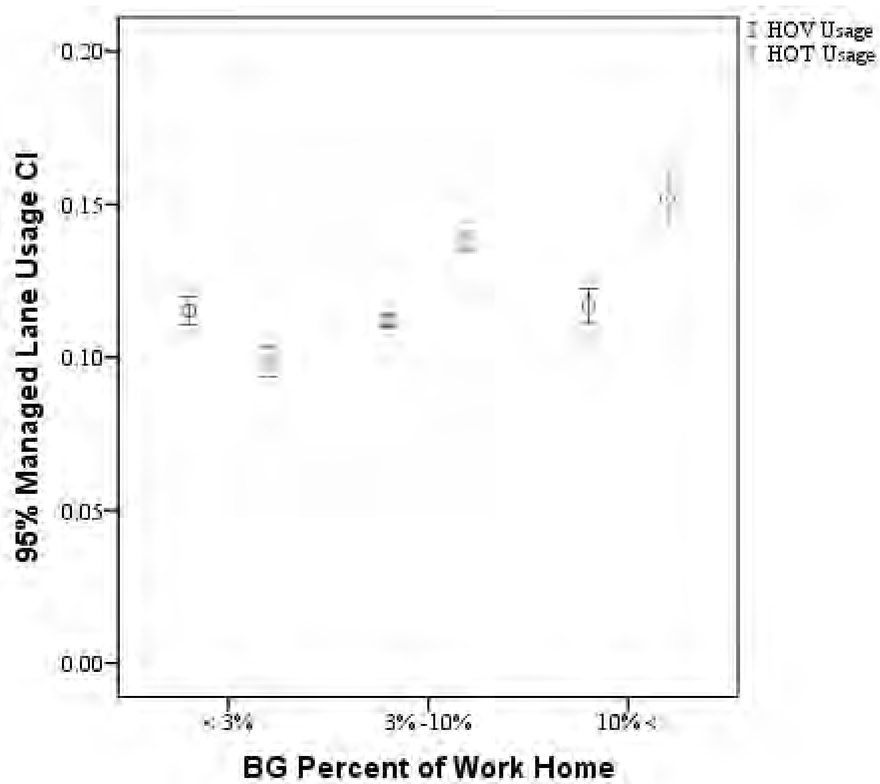


Figure 34 Block Groups Managed Lane Usage across Ratio of Work Home Categories (N=2,102)

5.4.2. Block group level Generalized Linear Modeling

5.4.2.1. Generalized Linear Models

The generalized linear model (GLM), which is a flexible generalization of ordinary linear regression that allows for response variables to have other than a normal distribution, is employed in the subsequent analyses. The response variables are managed lane usage rates between zero and one (π) which prohibits the use of ordinary least square. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value. For binary data (each license plate observation is either from the managed lane or general purpose lanes), the link function maps from $0<\pi_i<1$ to $\eta_i \in R$, and two links are commonly used: (1) Logit

(Equation 1), (2) Probit (Equation 2); where ψ (·) is the normal cumulative distribution function (MacCullagh and Nelder, 1989). In this study, Logit and Probit link functions have been employed and compared using goodness of fit parameters.

$$\boldsymbol{\eta_i = \log\left(\frac{\pi_i}{1-\pi_i}\right)} \qquad \textbf{Equation 1}$$

$$\boldsymbol{\eta_i = \psi^{-1}\,(\pi_i)} \qquad \textbf{Equation 2}$$

Because the model employs link function, the logistic regression coefficients are not easy to interpret. Instead, we have to translate using the exponent function. The odds ratio is equal to Exp (B), is a measure of association between an exposure and an outcome. The OR (odds ratio) represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure. When a logistic regression is calculated, the regression coefficient (B) is the estimated increase in the odds of the outcome per unit increase in the value of the exposure. In other words, the exponential function of the regression coefficient (Exp (B)) is the OR associated with a one-unit increase in the exposure (Szumilas, 2010).

In practice, when there is a positive relationship between a predictor and an outcome (B > 0), the OR is greater than 1 and as the B increases the OR increases. The interpretation is that when the scale predictor increases by one unit, the probability that the outcome happens (vs. the other alternative happens) increases by a factor of OR. Similarly, if there is a negative relationship between a predictor and an outcome (B < 0), OR is less than 1and as the B decreases the OR becomes closer to the zero. For non-scale predictors, the odds ratio will be interpreted in comparison. For example, if the predictor has two categories (male vs. female) the OR of one category will be assumed to be one (for example: male), while the OR of the other category (in this case: female) will be calculated. If the calculated OR (female) is more than 1, it implies that it is OR times more probable that the outcomes happens if the predictor is female.

With logistic regression, instead of R^2, there are other indicators for goodness of fit. The ρ^2 measures how much the log likelihood of the fitted model improved compared to the null model (Equation 3). In logistic regression deviance is analogous to the sum of squares in linear regression and is a measure of lack of fit to the data (J. Cohen and P. Cohen, 1975). Deviance is calculated by comparing a given model with the saturated model – a model with a theoretically perfect fit (Equation 4). The Pseudo R^2 shows the percentages of improvement in the model fit (smaller deviance), by comparing the deviance of the fitted model to the deviance of the null model (Equation 5). Therefore, as the model fit improves the deviance should decrease and the Pseudo R^2 becomes closer to 1. Normally, the ρ^2 and $Pseudo\ R^2$ goodness of fit measures are very close but not equal. The last goodness of fit measure, AIC, is illustrated in Equation 6; the k stands for number of parameters in the model. The smaller AIC indicates better goodness of fit.

$$\boldsymbol{\rho^2 = \frac{\text{LL (Null Model)} - \text{LL(Fitted Model)}}{\text{LL (Fitted Model)}}}$$ **Equation 3**

$$\textbf{Deviance (Fitted Model)} = -2\ln\frac{\textbf{LL (Fitted Model)}}{\textbf{LL (Saturated Model)}}$$ **Equation 4**

$$\textbf{Pseudo } \mathbf{R^2} = \frac{\textbf{Deviance (Null Model)} - \textbf{Deviance (Fitted Model)}}{\textbf{Deviance (Null Model)}}$$ **Equation 5**

$$\mathbf{AIC = 2k - 2\ln(LL)}$$ **Equation 6**

Lastly, the Omni test examines the hypothesis of whether the built model is significantly better than the constant only model in predicting the response variables. In the case of this section, all the models are significant with p-value less than 0.001 at 95% confidence.

5.4.2.2. Model Establishment

The dependent variable can be defined as total number of times that a license plate from a block group was observed along managed lane, normalized by total number of times the license plate was observed across all lanes from the same block group. The ratio of HOV lane frequency of observation to the corridor frequency of observation provides "HOV Usage" for each block group. Similarly, "HOT Usage" ratio has been calculated. "HOV Usage" and "HOT Usage" have been illustrated in Figure 35 in five quantiles across the corridor commutershed block groups. Considering average frequency of 600 observations per block group, block groups with fewer than 100 observations were removed from the graphical representation and statistical analysis.

Comparing HOV lane usage to HOT lane usage spatial distribution, the block groups belong to the top quantile of HOV usage are more concentrated in far north of the corridor, while block groups belong to the top quantile of HOT usage are more concentrated in very close north of the corridor (very dark brown block groups).

The revealed impact of congestion pricing on the spatial distribution of the commutershed is somewhat counter-intuitive. Because HOT lane provides free flow speeds for all the users who are willing to pay toll, one might expect to observe expansion of the commutershed after managed lane implementation. However, the managed lane commutershed retracted after the HOV to HOT conversion. While this section looks at the underlying socioeconomic relationships, further spatial analysis is conducted and presented in Chapter 8 to help explain the potential causes of this impact.

Considering the significantly different spatial distribution of the HOV and HOT lane usage, and practical aspects associated with access to the managed lane, block group location is considered as a predictor variable. A dummy variable (factor predictor) has been employed to divide the block groups in the study to three groups based on their location in the map. A small map in the left top corner of HOT map in Figure 35 illustrates these three spatial groups. The location categorical variable is zero for the blue block groups (south-west of the corridor), is one for the yellow block groups (North-east of the corridor), and is two for the red block groups (Farther North-east of the corridor). SPSS treated this location as two dummy variables for three location categories.

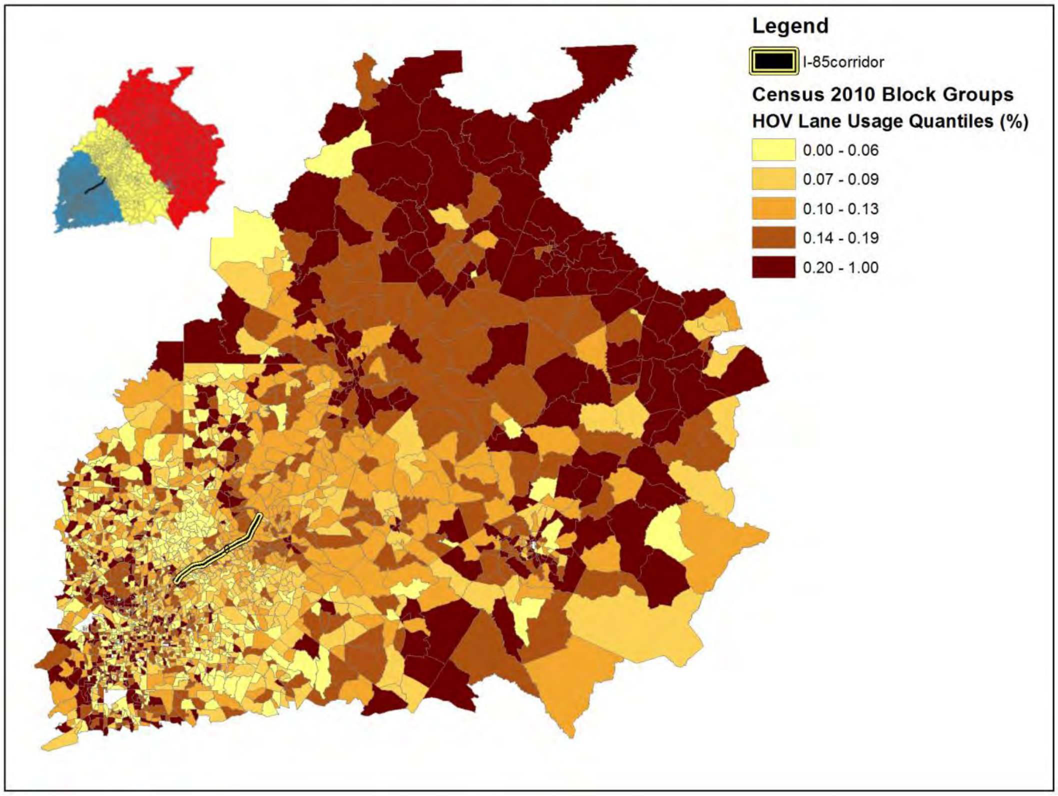
The three areas have been selected to coordinate with the commutershed spatial distribution HOT and COLD spots for both HOV and HOT lanes. Area zero has little practical use of the managed lane for commute travel because the user must move away from the downtown to access the lane. The reason that real distance value has not been used is for macroscopic consideration of both distance and direction with respect to the corridor. It is expected that location two becomes significant with larger power for "HOV Usage" while location one becomes significant with larger power for "HOT Usage" considering the spatial distribution maps.

The other independent covariates are socioeconomic variables at the block groups level, including: block group average household size ("HHSize"), block group median household income ("Income"), block group average individuals' age ("Age"), block group average household vehicle ownership ("Vehicles"), block group percent of people with masters or professional degree or doctorate degrees ("Education_high"), block group percent of female individuals ("Female"), block group average commute travel time ("Travel Time"), block group percent of households drive to work ("Drive"), and block group percent of individuals work at home ("Work_home").

Table 10 tabulates means and standard deviations of the covariate predictors. Because, the relation between income and travel behavior attributes are usually nonlinear (Kitamura, 1997) the square of block group median household income has also been considered as a predictor ("Income_sq"). To simplify names of the variables, the acronyms mentioned in the parenthesis will be used in the remaining parts of this section.

Table 10 List of Block group Level Predictors

Predictor	Mean	Std. Deviation
Household Size	2.6	0.5
Annual Income	$ 61,956	$ 34,981
Age	35.9 yrs	5.7
Vehicle Ownership	1.8	0.5
High Education	11%	0.07
Female	51%	0.06
Travel Time	31 min	7
Drive to Work	85%	0.12
Work at Home	5%	0.05



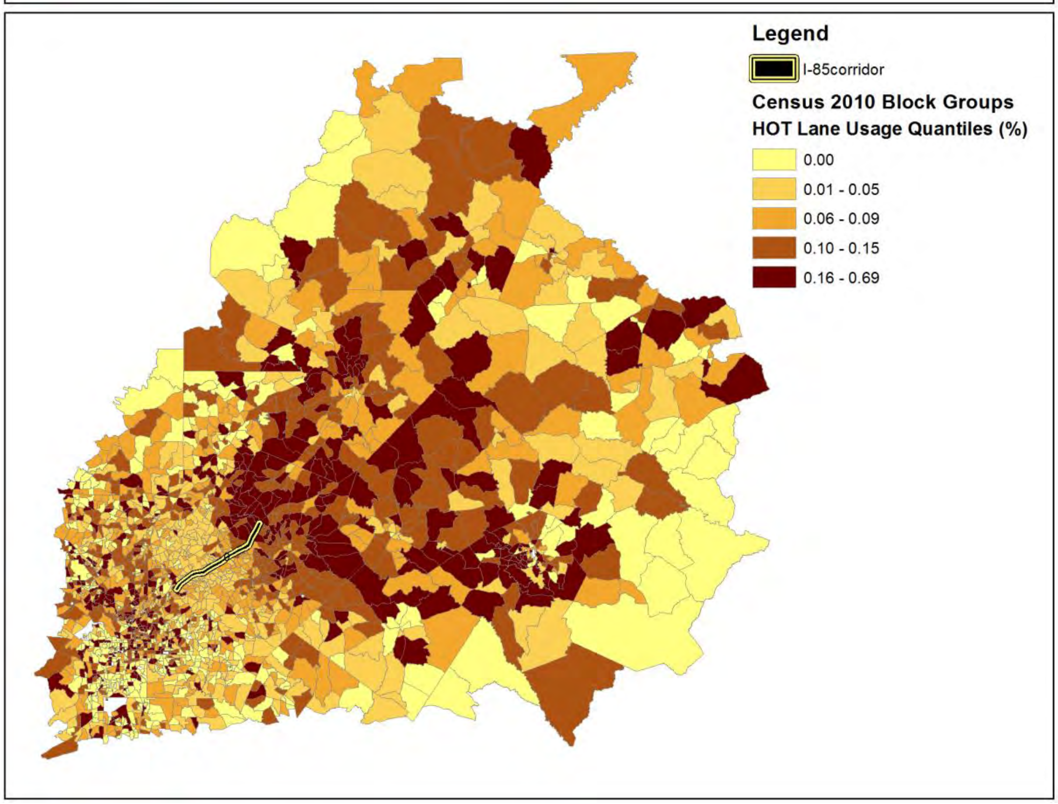


Figure 35 Managed lane usage quantile maps (Top: HOV Usage, Bottom: HOT Usage)

5.4.2.3. Results

Because different block groups have different frequency of observation and therefore different effects on the commutershed, “Weights” have been employed in the analysis. Block groups with smaller sizes and farther distance to the corridor have lower observation frequency and should have smaller weights. Block group’s total frequency of observation divided by average frequency across all the block groups have been computed as analysis weights. This is a standard method of weight calculation in Statistical packages (in this case: SPSS) which keeps the scale of the analysis similar to before weighting implementation. That is to say, while the number of block groups in the study equals 2,102, the sum of weight variable also equals 2,102.

In addition to the aforementioned block groups’ socioeconomic variables (ten variables) and location factor, the Logit transformation of the “HOV Usage” variable has also been considered as a predictor in the model (“Logit (HOV)”). The integration of historic HOV usage in HOT decision making serves to enhance the model, but is not the dominant factor. The outputs of the “HOT Usage” GLM models have been summarized in Table 11 and the outputs of the “HOV Usage” GLM models have been summarized in Table 12. Most of the coefficients are significant (bold font).

One difference across the developed GLM models are the applied link functions. While both Logistic and Probit link functions have been employed, Binary Probit models demonstrate a better fit for “HOT Usage”. Therefore only Binary Probit models have been developed for “HOV Usage” and tabulated in Table 12. The other difference is consideration of Logit (HOV) and location factor as predictors in the model. Location factor cannot easily be replicated in any future studies. For example, the location dummy variable can easily be replicated for Northeast corridor (considering similar commutershed spatial distribution); but, it cannot easily be replicated for I-285 considering totally different commutershed spatial distribution. Similarly, HOV lane use frequency would not be always available in future studies; because there may not be any HOV lane existing or no data would have been collected. Since these two variables could not easily be obtained in any future replication of this study, models only based on block group socioeconomic attributes (publicly available free of charge) have also been presented. Lastly, if the best model (in terms of goodness of fit) has any insignificant variable, the final model, excluding those insignificant variables, has also been presented.

The parameters estimates have all been displayed in the tables (the insignificant coefficients are not bold). Comparing six models computed for "HOT Usage", model 1, 2, and 3 are the similar models to 4, 5, and 6 respectively with the only difference of using Logistic link instead of Probit link (Table 11). Comparing the three pairs, Probit models have slightly better fit compared to the Logit models. Model 4 has the best fit by including all the 10 socioeconomic variable, location factor and Logit (HOV) with $\rho^2(c)$ of 0.455. All the variables are significant except the location factor equal to two (far north-east group). Model 5 is exactly same as model 4 with exclusion of Logit (HOV) variable. While all the parameter estimates remain significant and stable, the $\rho^2(c)$ decreased to 0.390. Model 5 is useful in future in cases without an HOV lane or any pre-conversion information about HOV lane use. Because any future corridor may not have same geographic situation as I-85 the same model without the location factor and Logit (HOV) parameter have also been developed as model 6. The model 6 $\rho^2(c)$ decreased to

0.325 and “Work_home” variable becomes insignificant. The AIC parameter changes are in accordance with $\rho^2(c)$.

The goodness of fit of the “HOV Usage” models is significantly lower than the “HOT Usage” models. The $\rho^2(c)$ is roughly 0.4 for “HOT Usage” models (even without the use of historic HOV usage as an input variable), but only 0.04 for “HOV Usage”. This was expected since HOV lane usage is dependent on the ability to carpool while HOT lane usage depends more on the ability to pay toll and value of time. Although the model fit for “HOV Usage” is not satisfying, all the models and most of the variables are significant. Model 1 is the full model which includes all the ten socioeconomic variables and location factor as model predictors. All the variables except “Drive”, Work_home” and location variable when it is equal to one (north-eat group) are significant. Model 2 is same as model 1 with exclusion of location factors. The $\rho^2(c)$ only decreased from 0.039 in model 1 to 0.033 in model 2 by removing the location parameter. All the other variables are significant except for “Drive” and “Work_home” variables same as model 1. Model 3 (final model) excludes the insignificant variables and has the same $\rho^2(c)$ as model 1.

Table 11 HOT lane Usage Block Group Level Model Output

		Model 1	Model 2	Model 3	Model 4 (Final)	Model 5	Model 6
	Model Type	Binary Logistic	Binary Logistic	Binary Logistic	Binary Probit	Binary Probit	Binary Probit
Predictors	**Intercept**	**-1.635**	**-2.452**	**-3.145**	**-0.965 (0.381)***	**-1.441**	**-1.831**
	HH Size	**-0.203**	**-0.201**	**-0.199**	**-0.110 (0.895)**	**-0.108**	**-0.108**
	Income	**2.55E-5**	**2.35E-5**	**2.97E-5**	**1.35E-5 (1.000)**	**1.25E-5**	**1.59E-5**
	Income Square	**-9.62E-11**	**-8.72E-11**	**-1.18E-10**	**-4.99E-11 (1.000)**	**-4.58E-11**	**-6.30E-11**
	Age	**-0.014**	**-0.011**	**-0.015**	**-0.008 (0.992)**	**-0.006**	**-0.008**
	Vehicle Ownership	**0.117**	**0.097**	**0.135**	**0.065 (1.067)**	**0.054**	**0.073**
	High Education	**-0.578**	**-0.116**	**-0.498**	**-0.310 (0.733)**	**-0.056**	**-0.273**
	Female	**1.144**	**0.858**	**0.847**	**0.627 (1.873)**	**0.477**	**0.469**
	Travel Time	**0.002**	**0.004**	**0.008**	**0.001 (1.001)**	**0.002**	**0.004**
	Drive to Work	**-0.568**	**-0.663**	-0.07	**-0.294 (0.745)**	**-0.337**	0.005
	Work at Home	**1.146**	**1.048**	**2.059**	**0.651 (1.917)**	**0.614**	**1.174**
	Location = 2	-0.024	**0.164**		0.009 (0.991)	**0.092**	
	Location = 1	**0.235**	**0.240**		**0.129 (1.138)**	**0.130**	
	Logit (HOV)	**0.452**			**0.250 (1.284)**		
Degree of Freedom		1,896	1,896	1,896	1,896	1,896	1,896
LL		-18,899	-21,068	-22,337	-18,732	-20,971	-22,264
LL (c)		-34,396	-34,396	-34,396	-34,396	-34,396	-34,396
ρ^2 (c)		0.450	0.387	0.351	0.455	0.390	0.352
AIC		37,829	42,163	44,696	37,493	41,969	44,550
Omni Test Sig.		0.000	0.000	0.000	0.000	0.000	0.000

*The numbers in parenthesis are Odds Ratios

Table 12 HOV Lane Usage Block Group Level Model Output

		Model 1	Model2	Model 3 (Final)
	Model Type	Binary Probit	Binary Probit	Binary Probit
Predictors	**Intercept**	**-1.232**	**-1.236**	**-1.251 (0.285)***
	HH Size	**0.025**	**0.025**	**0.025 (1.018)**
	Income	**-2.079E-6**	**-2.424E-6**	**-2.125E-6 (1.000)**
	Income Square	**8.745E-12**	**1.029E-11**	**9.141E-12 (1.000)**
	Age	**0.006**	**0.006**	**0.006 (1.006)**
	Vehicle Ownership	**-0.047**	**-0.043**	**-0.048 (0.954)**
	High Education	**0.486**	**0.491**	**0.491 (1.635)**
	Female	**-0.287**	**-0.293**	**-0.293 (0.746)**
	Travel Time	**0.002**	**0.002**	**0.002 (1.002)**
	Drive to Work	-0.025	-0.019	
	Work at Home	0.030	0.043	
	Location = 2	**0.222**		**0.223 (1.250)**
	Location = 1	-0.001		-0.001 (0.999)
Degree of Freedom		1,896	1,896	1,896
LL		-21,951	-22,097	-21,959
LL (c)		-22,852	-22,852	-22,852
ρ^2 (c)		0.039	0.033	0.039
AIC		43,937	44,217	43,940
Omni Test Sig.		0.000	0.000	0.000

*The numbers in parenthesis are Odds Ratios

Investigating the parameter estimates could also envision the relationship between socioeconomic attributes and managed lane usage behavior. One main difference between HOV and HOT lane models are the location factor parameter estimates. Location two which is far north-east of the corridor is significant for "HOV Usage" models while location one which is near north-east of the corridor is significant for "HOT Usage" models. Therefore the odds of using HOT lane is 14% higher from closer north areas to the corridor while the odds of using HOV lane is 25% higher from farther north areas to the corridor. This finding corresponds with what we have observed in Figure 35.

In terms of household size, the odds of HOT usage decreases 11% and the odds of HOV usage increases 3% as the blockgroup average household size increases by one unit. This difference comes from the larger probability of carpooling and using HOV lane in larger households (fampooling). Although, the HOT lane provides free trip for HOV3+ carpoolers, only about 6% of the HOT lane trips fall in this category.

With respect to income and square of income variables, both models show significant nonlinear relationships with different parameters. The application of both income and income square in a model has been initially proposed by Kitamura et al., 1997. The odds ratios are very close to one because income and income square units are dollars and the impact of one dollar increase/decrease in income is insignificant. However, the odds values, for $10,000 change in income, are meaningful. The odds of HOT Usage increases 15% with $10,000 increase of block group median household income up to around $130,000, while odds of HOV Usage decreases 2% with $10,000 increase of block group median household income up to $94,000. These findings are intuitive to the nature of managed lane. While high income households may be less willing to carpool, they may be more willing to pay toll and get benefit of HOT lane.

Although the odds ratio is very close to one for age variable (impact of one year change in block group average age), the odds values for 10 years change in age are meaningful. The odds of HOT usage decreases 7% and the odds of HOV usage increases 6%, as the average block group age increases by 10 years. This result implies that there may be higher probability of carpooling by older users and higher chance of paying tolls by younger users. It should be considered that age variable is also correlated with presence of children, which also potentially increase the chance of carpooling.

In terms of vehicle ownership, the odds of HOT usage increases by 7% and the odds of HOV usage decreases by 5%, as the average household vehicle ownership per block group increases by one unit. This again intuitively coordinates with the concept of paying toll versus carpooling. Having lower number of vehicles likely increases the chance of carpooling, while a higher number of vehicles per household likely decreases the chance of carpooling and increases the chance of using HOT lane.

Because well educated people have a higher value of time, higher HOT lane use is expected. However, the odds of HOT usage decreases 3%, and the odds of HOV usage increases 5%, as the percent of high educated people (Masters or more) in a block group increases by 10% (the 10^{th} root of odds ratios are presented, because the predictors are percent values). While this finding seems counter-intuitive, the real impact of education variable may be embedded in other variables such as income, job type, and location. For example, the penalty for delay in jobs with high education requirement is usually much smaller compared to jobs with low education requirement.

In terms of gender, the odds of HOT usage increases 6% and the odds of HOV usage decreases 3%, as the percent of female population in the block group increases by 10%. This finding matches with previous study which expressed higher number of toll trips for women compared to men considering the higher levels of household responsibilities that they also often undertake (Devarasetty, 2012).

Household commute travel time impact is very small and practically insignificant. The odds of HOT usage increases 1% and the odds of HOV usage increases 2%, as the average commute travel time increases 10 min per block group. "Drive" variable which indicates percent of households who drive to work was not significant for HOV model; however, the odds of HOT usage decreases 3% as the percent of people who drive to work increases by 10%. This finding is counter-intuitive, because it is expected that using the drive mode for commute increase the chance of using HOT lane. However, because of the drive-oriented commute style of Atlanta, other modes may be misrepresented.

The last variable is "Work_home" which indicates percent of people who work from home is again not significant for HOV model while it is significant and positive for HOT model. The odds of HOT usage increases by 7% as the percent of home workers in a block group increases by 10%. This is intuitive, because people who are working at home should care more about the value of time, because they should be productive at their own time such as consultants.

5.4.3. Discussion

Although aggregate data are publicly available at no cost, the aggregation functions removed the desired level of variation across the cases since a single central value is assigned to all the households in a certain boundary. However, aggregated data could still show the trends among population groups significantly. Most of the parameters impacts on managed lane usage coordinate well with previous literature findings.

The presence of potential correlation between some variable may be a potential reason for these few counter-intuitive results. For example, a portion of "Education" impact in the model may be embedded in other variables such as "Income". Further research could possibly help understand the roll of each variable in the model independently. Of course, supplementary analysis at household level on the large-scale collected data could improve our knowledge regarding the socioeconomic impact of the managed lanes.

Deploying the developed model, future managed lane traffic and revenue studies could import the socioeconomic characteristics of the block groups in the corridor commutershed and predict an initial estimate of users' response to pricing. Hence, findings of this study could assist regional planners deciding about the priority and feasibility of congestion pricing projects while advancing the socioeconomic knowledge of congestion pricing.

CHAPTER 6

MARKET ANALYSIS[5]

This chapter assesses traveler response to the HOV-to-HOT conversion within four user markets: general purpose lanes users that remained in the general purpose lanes after the conversion (the GP market), the HOV lane users that shifted to the general purpose lanes (the HOV market), general purpose lanes users that shifted to the HOT lane (the HOT market), and HOV lane users that are using the HOT lane (the HOV/HOT market). The markets are defined based upon households' lane use frequency before and after the conversion.

The significance of differences in household level attributes (based on marketing data) of such as income, vehicle ownership, household size, ethnicity, age, gender, income and education across the markets are examined. The major strength of this analysis compared to other sections is the utilization of three different socioeconomic data sources simultaneously to examine the socioeconomic aspects of a pricing scenario.

6.1. Markets Establishment

Cluster analysis has been used to divide the I-85 users into four exclusive groups based on the percent of trips been observed in the managed lane. The 13,476 plates that were observed a minimum of 20 times, considered as frequent commuters, are used in this analysis (5.6% of the households representing 42% of the trips).

---

[5] This section has been published: S. Khoeini, R. Guensler. "HOV to HOT conversion Socioeconomic Assessment: Atlanta I-85 HOV-to-HOT conversion"; Transportation Research Board, 2014.

Based on managed lane use frequencies, two variables illustrating the household managed lane usage have been created: "*HOV usage*" is defined as the percent of the observations for a vehicle in the HOV lane before the conversion, and "*HOT usage*" is the percent of observations for a vehicle in HOT lane after the conversion.

The "*HOV usage*" and "*HOT usage*" parameters are then used to assign the observed households into four market clusters: GP (General Purpose) Market (commute via HOV 20% or less before, and commute via HOT 20% or less after), *HOV Market* (commute via HOV more than 20% before, but commute via HOT 20% or less after), *HOT Market* (commute via HOV 20% or less before, but commute via HOT more than 20% after), *HOV/HOT* Market (commute via HOV more than 20% before, and commute via HOT more than 20%). The 20% value was obtained from a Two-Step Cluster analysis method in SPSS which employs an agglomerative hierarchical clustering method (Kaufman and Rousseeuw, 2009).

Table 13 illustrates the users cluster analysis and Figure 36 Clustering Distribution illustrates clusters distributions. The GP market consists of 66% of corridor users (the majority of the users), who used the HOV lane, on average, 2% of time and now use the HOT lane only 1% of time. The GP market is not significantly impacted by the conversion (considering no change in general purpose lanes traffic condition). Households in the HOV market used the HOV lane, on average, 48% of the time, but now use the HOT lane, on average, only 1% of time. The HOV market represents 9% of total users (in the sample), and its members were potentially negatively impacted by the conversion as they are no longer using what should be the faster lane. Households in the HOT market used the HOV lane, on average, only 3% of the time, but now use the HOT lane, on average, 59% of the time. The HOT market consists of 18% of users, and these users are positively impacted by the conversion. The HOV/HOT market, who used the HOV lane, on average, 56% of time, continue to use HOT lane, on average, 64% of time. The HOV/HOT market consists of only 7% of the users, and these users are also positively impacted by the conversion.

Table 13 Users Cluster Analysis

Cluster	Average HOV Usage	Average HOT Usage	HH Sample Size
GP Market	2%	1%	8,877
HOT Market	3%	59%	2,487
HOV Market	48%	1%	1,191
HOV/HOT Market	56%	64%	921

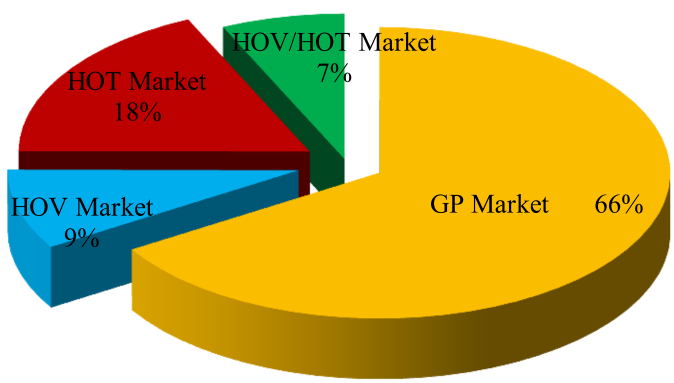


Figure 36 Clustering Distribution

The same cluster analysis is similarly conducted for the households that participated in the Volpe two-day travel survey. However, the frequency of trips along I-85 from the survey is very low (only four days of data collection) compared to the license plate data. The same cluster analysis rules were used to divide the 1,256 frequent corridor users (at least one household trip per day along I-85) in the Volpe sample into the four market clusters. Cluster analytical results were obtained as follows: GP market: 62% (average HOV usage: 2%, average HOT usage: 1%), HOV market: 8% (average HOV usage: 57%, average HOT usage: 1%), HOT market: 16% (average HOV usage: 2%, average HOT usage: 70%), and HOV/HOT market: 14% (average HOV usage: 81%, average HOT usage: 85%).

The sample fractions by market (percentage of households in each market cluster) are similar across license plate data and survey data; however, there is a fairly large difference associated with the HOV/HOT market. HOV/HOT market accounts for only 7% of households in the license plate data but accounts for 14% of households in the survey data. This outcome may not be surprising considering that survey recruitment was conducted before conversion with a higher weighting toward the HOV lane, and perhaps there is a slight retention or self-selection bias toward HOV/HOT users in the Volpe survey.

6.2. Socioeconomic Attributes across Markets

Among various socioeconomic attributes, this section concentrates on the ones more applicable to travel behavior studies, including: household annual income, vehicle ownership, household size, and ethnicity, for comparative analysis across data sources (aggregated American Community Survey data, Volpe household travel survey, and marketing data). The demographic profiles of other socioeconomic attributes (using marketing data) have also been presented across the four clusters (GP market, HOV market, HOT market, and HOV/HOT market).

Whereas the frequencies of total observed trips across the households are not equal, the analytical results are weighted accordingly. Following SPSS analysis guidelines, for each household, the sample weight is equal to number of trips per households divided by average number of trips by all the households. Lastly, the statistical tests significance is 0.05, corresponding to 95% confidence. Given the large sample size employed, even small differences in analytical results are often statistically significant (but may not be that meaningful when such differences are small).

6.2.1. Household Annual Income

Household annual income is one of the main attributes in socioeconomic impact assessments, especially in environmental justice studies. Table 14 presents the income statistics including 5% and 95% quantiles, 50% inter-quantile range, median, mean and mean confidence intervals.

Using the marketing data, average income for the GP market ($71,030) is significantly higher than the HOV market ($67,162); however, the magnitude of the difference is fairly small (5%). Both the HOV and GP market average income are significantly lower than the HOV/HOT market ($77,801), and the HOT market ($81,263) average income which are not significantly different. The largest difference is between HOV and HOT market ($14,101), which is 21% of the HOV market average income.

Median of Census block group household annual income has been obtained from ACS for each household. Using ACS data, the HOV market has the lowest average income ($72,475) followed by GP market ($73,210), followed by the HOV/HOT market ($76,093), and then the HOT market ($78,353). As before, the largest difference ($5,878) is between HOV and HOT market, which is only 8% of the HOV market average income. ACS data show the same order of the four markets as household level marketing data; however, the differences are significantly smaller.

Based on Volpe travel diary survey, the GP market exhibits the lowest average income ($85,764) and the HOT and the HOV/HOT markets ($98,973 and $99,678 respectively) exhibit very close and the highest average income. However, the HOV market mean income is not significantly different from any of the clusters; this was expected given the low sample size (100 out of 1,256 survey respondents belong to HOV market). The largest difference ($13,914), between GP and HOV/HOT market, corresponds to 16% of the GP market average income.

The significantly smaller inter-quantile range in the ACS data, compared to the marketing and survey data, results from the ACS aggregation process, which essentially removes variability and brings values closer to the average. Indeed, the smaller scale of distinction between higher and lower households attributes is the main disadvantage of the ACS data. There is no way to know for sure which data source contains data that are closest to reality for each household. However, if the ACS data are closest to actual household values, the Volpe household survey data are likely overestimating the households' income (perhaps through sample bias) and marketing data may be closer to reality.

While the median values of marketing data are different from the mean values (primarily lower because of the long right-hand tail for income distributions), similar trends in differences are also observed across the markets. The median values of ACS data are the same as the mean values (means/medians of the means are normally distributed based on central limit theorem). On the contrary, all the median values from survey data are roughly equal.

Table 14 Household Income Statistics across Markets for Each Data Source

		GP Market	HOV Market	HOT Market	HOV/HOT Market
Marketing (n= 17,857)	5% Quantile	$25,000	$25,000	$25,000	$25,000
	95% Quantile	$112,500	$112,500	$137,500	$137,500
	IQ Range	$42,500	$42,500	$50,000	$67,500
	Median	$62,500	$62,500	$62,500	$62,500
	Mean	$71,030	$67,162	$81,263	$77,801
	Mean CI LB	$70,256	$65,011	$79,695	$75,178
	Mean CI HB	$71,805	$69,313	$82,830	$80,425
ACS (n= 17,857)	5% Quantile	$37,281	$35,781	$42,237	$40,139
	95% Quantile	$107,940	$109,750	$112,579	$117,292
	IQ Range	$29,804	$31,524	$24,985	$28,233
	Median	$71,918	$70,951	$75,659	$73,409
	Mean	$73,210	$72,475	$78,353	$76,093
	Mean CI LB	$73,040	$71,983	$78,052	$75,573
	Mean CI HB	$73,381	$72,968	$78,655	$76,613
Survey (n= 1,256)	5% Quantile	$17,500	$17,500	$17,500	$17,500
	95% Quantile	$175,000	$194,023	$225,000	$175,000
	IQ Range	$62,500	$62,500	$62,500	$62,500
	Median	$87,500	$87,500	$87,500	$87,500
	Mean	$85,764	$92,650	$98,973	$99,678
	Mean CI LB	$82,274	$82,639	$91,056	$91,676
	Mean CI HB	$89,254	$102,661	$106,890	$107,681

To improve our understanding about the economic impact on user lane choice, Figure 37 illustrates the distribution of income groups across the four markets. Because of the substantial difference between the income distribution of ACS block group level and the household level data sources, only survey and marketing data sources are presented. The users are divided to four income groups: very-low-income (less than $30,000), low-income (between $30,000 and $60,000), medium-income (between $60,000 and $110,000), high-income (more than $110,000).

The main purpose of illustrating survey data adjacent to marketing data is to show that marketing data exhibit similar trends to survey data. However, small sample size of survey data prohibits any statistically valid conclusion across income groups. For example, only 8 households represent the very-low-income category of the HOV market. Furthermore, Pearson Chi-Square test doesn't show significant difference in income group distributions across the markets for the survey data (p-value: 0.265). However, the same test shows a large significant difference for the marketing data (p-value: 0.0000). Therefore, detailed explanation of the income variations across the markets, provided below are based on the marketing data.

Generally, the HOV and GP markets present similar distributions and the HOT and HOV/HOT markets present almost exactly the same distributions. Based upon marketing data, the very-low-income group usage proportions are very close across the markets (from 4% to 5%) indicating that there does not appear to be a major impact on very-low-income commuters. Whereas, low-income users account for 27% and 28% of the HOV and GP markets respectively, they account for 22% of the HOT and HOV/HOT markets. Indeed, the GP and HOV markets include 22% more low-income households compared to the HOT and HOV/HOT markets, implying a significant lower use of HOT lane by low-income users.

Whereas, medium income users account for 37% and 41% of the GP and HOV markets respectively, they account for 33% of the HOT and HOV/HOT markets. Therefore, the GP market includes 12% more medium income users compared to HOT and HOV/HOT market. Lastly, high-income users account for 27% and 30% of the HOV and GP markets respectively, versus 40% and 41% of the HOV/HOT and HOT markets.

Indeed, the HOV/HOT and HOT markets include 33% more high-income users compared to GP market.

In summary, proportions of all income groups are larger across the GP and HOV market compared to the HOT and HOV/HOT markets, except for high-income users. Particularly, the HOT and HOV/HOT markets contain high-income users (more than $110,000 household annual income) almost 50% more than the very-low-income and low-income users (less than $60,000 household annual income). However, the GP and HOV markets contain the high-income users (more than $110,000 household annual income) almost 10% less than very-low-income and low-income users (less than $60,000 household annual income). Although there are variations across markets and income groups' categorical distributions, all the four income groups are well distributed across the four markets illustrating the effectiveness of the conversion for all users.

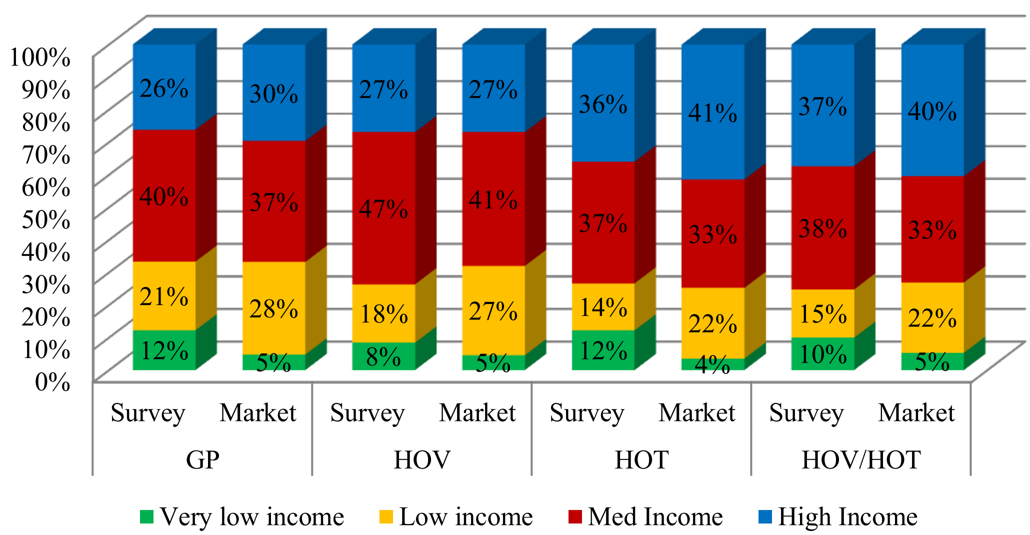


Figure 37 Household Income Categorical Distribution across Markets for Each Data Source

6.2.2. Vehicle Ownership

Vehicle ownership is a key attribute in transportation studies such as travel demand model development and air quality impact assessment. Household vehicle ownership is not available as a separate variable in ACS data at the Block Group level. The marketing data that were purchased did not provide vehicle ownership either. In this section, vehicle ownership is obtained from the vehicle registration database and Volpe household survey.

Table 15 presents the vehicle ownership statistics including 5% and 95% quantiles, inter-quantile range, median, mean and mean confidence intervals. Registration data illustrates that the HOV market has the significantly higher average vehicle ownership (3.18), followed by HOV/HOT market (2.98), which is significantly higher than HOT and GP market (2.84 and 2.83, respectively). The medians are equal to 3.0 for all markets except for the GP market which is 2.0.

Survey data indicate smaller vehicle ownership values (about 0.7 on average) across the markets compared to the registration database. This outcome was expected because many households are known to have registered vehicles that are not garaged at the registration location (Granell, 2002). For example, vehicles used by children that have moved to other locations (e.g., at college or in urban areas for new jobs) are often registered to parental households for insurance purposes. Many older vehicles may also be registered and garaged at the home location that are no longer actively used and therefore not reported in surveys.

Using survey data, similar trends of differences in average vehicle ownership are observed. HOV market indicates to have the highest number of vehicles (2.44) followed by HOV/HOT market (2.26), GP market (2.18) and HOT market (2.17). However, only the HOV market is statistically different from the GP and HOT markets. The medians are all equal to 2 across the markets. Registration data provide a wider distribution of vehicle ownership, with interquartile range of 2 and 95% quantile of 6 or 7, while survey data provide narrower distribution with interquartile range of 1 and 95% quantile of 4.

Table 15 Household Vehicle Ownership Statistics across Markets for Each Data Source

		GP Market	HOV Market	HOT Market	HOV/HOT Market
Registration Data (n= 16,033)	**5%**	1.00	1.00	1.00	1.00
	95%	6.00	6.00	7.00	6.00
	IQ Range	2.00	2.00	2.00	2.00
	Median	2.00	3.00	3.00	3.00
	Mean	2.83	3.18	2.84	2.98
	Mean CI LB	2.82	3.15	2.82	2.94
	Mean CI HB	2.84	3.22	2.86	3.01
Survey Data (n= 1,165)	**5%**	1.00	1.00	1.00	1.00
	95%	4.00	4.00	4.00	4.00
	IQ Range	1.00	1.00	1.00	1.00
	Median	2.00	2.00	2.00	2.00
	Mean	2.18	2.44	2.17	2.26
	Mean CI LB	2.11	2.25	2.04	2.11
	Mean CI HB	2.24	2.63	2.29	2.41

Figure 38 presents the distributions of vehicle ownership across the markets using the survey and the registration data. Unlike income distributions, vehicle ownership distributions are very different comparing the survey data to the registration-based data. The proportions of one and two vehicle ownership are larger based on the survey data, whereas the proportions of three, four or more vehicle ownership are larger based on the registration data. As mentioned earlier, the underlying reason for lower vehicle ownership reported through surveys is likely to be associated with vehicles registered at a home location that are garaged at different locations (Granell, 2002).

Given the small sample size of survey data, particularly after dividing households among 16 groups (four markets and four vehicle ownership groups), no statistically significant differences can be derived. Pearson Chi-square test doesn't show significant difference across the markets based upon survey data (p-value: 0.06). Hence, the registration data will be used for explaining the variations in vehicle ownership across the markets, keeping in mind that the significant difference (p-value: 0.0000); the household vehicle ownership values from this data source are likely to be somewhat higher than really exist.

The HOV market exhibits a substantially different distribution compared to the other markets. Surprisingly, 37% of the HOV market consists of users with that own four or more vehicles compared to between 27% and 30% of other markets. Accordingly, 29% of HOV market consists of owners of two vehicles compared to from 35% to 39% of other markets.

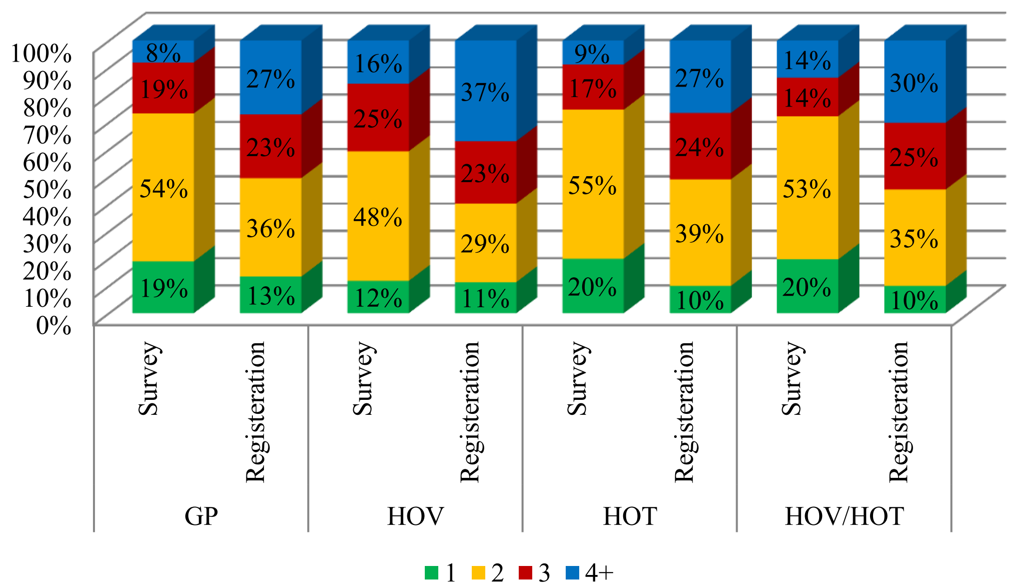


Figure 38 Household Vehicle Ownership Categories across Markets for Each Data Source

The HOV market consists of users who were taking advantage of the HOV lane before conversion, but are not using the HOT lane after conversion. One potential reason might be that HOV users preferred to carpool and use the HOV lane before, even though they could have driven alone (e.g., a fampool). After the conversion, they may no longer be carpooling or may not be willing to pay the toll. These findings matches well with our previous carpooling analysis which showed 44% of frequent HOV carpoolers have not been observed along the HOT lane after the conversion.

6.2.3. Household Size

Household size is another key attribute in transportation studies, particularly travel demand modeling. Table 16 presents the household size statistics including 5% and 95% quantiles, inter-quantile range, median, mean, and mean confidence intervals across the three data sources. Based upon the marketing data, the average household size is highest in the HOT market (2.97), followed by the HOV/HOT market and GP market (2.84 and 2.79, respectively, but not significantly different from each other), followed by HOV market (2.59). The median household size is two across all markets, except for the HOT market which is three.

By comparison, the survey data yield no significant differences across the markets, except for the HOV market which has significantly higher average household size (2.94) compared to the GP Market (2.58). In comparing the survey data to the marketing data, average household size values are very close for the HOT, HOV/HOT, and GP market groups. However, HOV market has the highest household size based upon survey data and the lowest household size based upon marketing data. The higher mean value for HOV market household size, using survey data, is not significant due to the small sample size (n=100).

ACS data exhibit very similar average household size values across the four markets. The reason again ties back to the usage of aggregate ACS data. The HOT market is the cluster with highest household size (3.10), followed by the HOV market (3.09), GP market (3.08) and HOV/HOT market (3.05), all of which are significantly different due to the large sample size.

With respect to data variation, marketing data exhibits the highest inter-quantile range of 3.0 more than survey data, ranging between one and two, and ACS data exhibits the lowest inter-quantile range of about 0.5 persons per household.

Table 16 Household Size Statistics across Markets for each Data Source

		GP Market	HOV Market	HOT Market	HOV/HOT Market
Marketing Data (n= 16,033)	**5% Quantile**	1.00	1.00	1.00	1.00
	95% Quantile	6.00	5.00	6.00	6.00
	IQ Range	3.00	3.00	3.00	3.00
	Median	3.00	2.00	3.00	3.00
	Mean	2.79	2.59	2.97	2.84
	Mean CI LB	2.78	2.55	2.95	2.80
	Mean CI HB	2.81	2.62	3.00	2.88
ACS Data (n= 16,033)	**5% Quantile**	2.35	2.40	2.42	2.27
	95% Quantile	3.63	3.63	3.60	3.61
	IQ Range	0.49	0.46	0.48	0.48
	Median	3.11	3.10	3.14	3.10
	Mean	3.08	3.09	3.10	3.05
	Mean CI LB	3.08	3.08	3.10	3.04
	Mean CI HB	3.08	3.10	3.11	3.06
Survey Data (n= 1,165)	**5% Quantile**	1.00	1.00	1.00	1.00
	95% Quantile	5.00	6.00	5.04	5.00
	IQ Range	1.00	2.00	2.00	2.00
	Median	2.00	3.00	2.00	2.00
	Mean	2.58	2.94	2.76	2.70
	Mean CI LB	2.50	2.67	2.58	2.51
	Mean CI HB	2.67	3.21	2.95	2.88

Figure 39 represents categorical distribution of household size across the marketing and survey data. ACS data are not presented because the household size values do not represent individual households and are obtained from aggregation functions. Like vehicle ownership, but unlike income, the distributions of household size appear to be different across the survey data and marketing data. The main source of dissimilarity comes from households with one and two members. Marketing data report more single person households whereas survey data report more two-person households. Meanwhile, the illustrated difference in household size across the markets, applying Pearson Chi-square test, is not significant based on survey data (p-value: 0.26), whereas it is significant based on the marketing data (p-value: 0.0000).

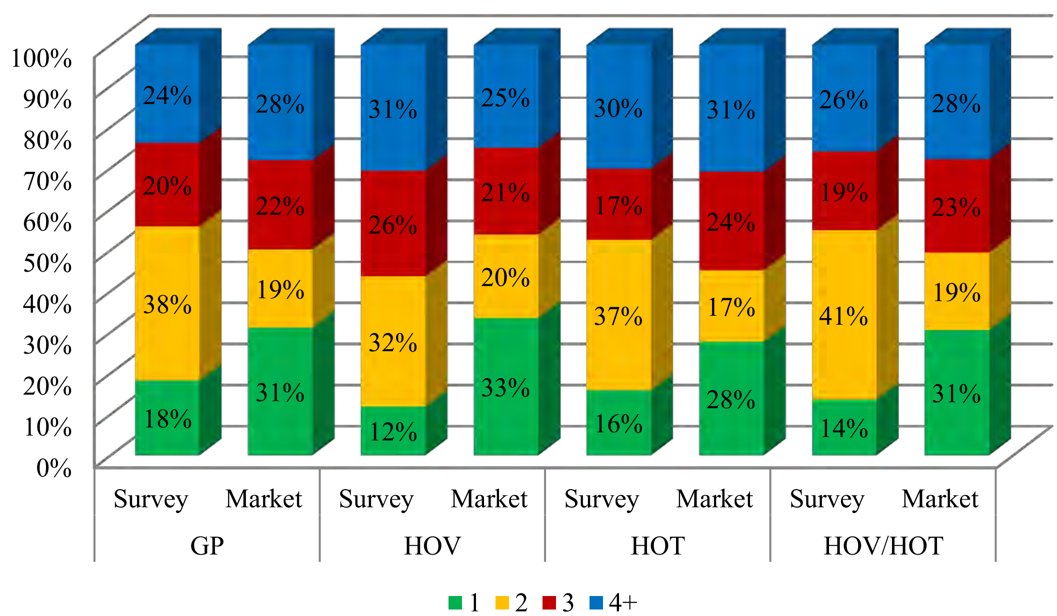


Figure 39 Household Size Categories across Markets for Each Data Source

Looking across the four user markets using the marketing data, the household size proportions are very close, indicating that household size is not as impactful as income and vehicle ownership. Because larger household size could potentially increase the chance of carpooling (fampooling), this outcome was counter-intuitive. Although the survey data exhibit significantly different results and insignificant distributions, they do not show as much difference in household size proportions across the markets, statistically. For example, HOV market sample size based on survey data is 100. Therefore, the portion of HOV market, indicating household size of one, has only 12 households which is substantially small to be the basis for a statistically significant conclusion.

6.2.4. Ethnicity

The last essential attribute for equity and environmental justice studies is ethnicity. Aggregate ACS data, which only provide the percentages of each ethnic group for each block group, have not been analyzed here. Figure 40 presents the categorical distribution of ethnic groups across the user markets based upon both marketing and survey data. Because ethnicity is a discrete variable, no central statistics could be derived similar to previous attributes and only Pearson Chi-square tests have been conducted. Pearson chi-square tests show no significant difference (p-value = 0.34) across the markets based on survey data, but shows a significant difference (p-value = 0.0000) across the markets, based on marketing data. In addition to significance of marketing data distributions, ethnicity is better quality marketing variable, because it does not change over time unless a household moves.

The HOV market illustrates extensive differences in ethnicity distribution compared to the other markets. The proportion of Hispanic population in the HOV market is about two times larger than in the GP and HOV/HOT markets and about three times larger than in the HOT market. Furthermore, the proportion of Asian population in the HOV market is about two times larger than in the HOT and GP markets. Accordingly, the proportion of White population is highest at HOT market (83%), and lowest at the HOV market (60%). Not much difference in proportion of African-American population was observed across the markets.

It is important to keep in mind that ethnicity is very well correlated with other socioeconomic attribute such as income, number of children, and home ownership. In addition, there is a great deal of spatial correlation to home locations (and perhaps work locations) as well. Hence, the difference in lane usage by ethnic groups may better reflect differences across the correlated variables rather than the specific ethnicity.

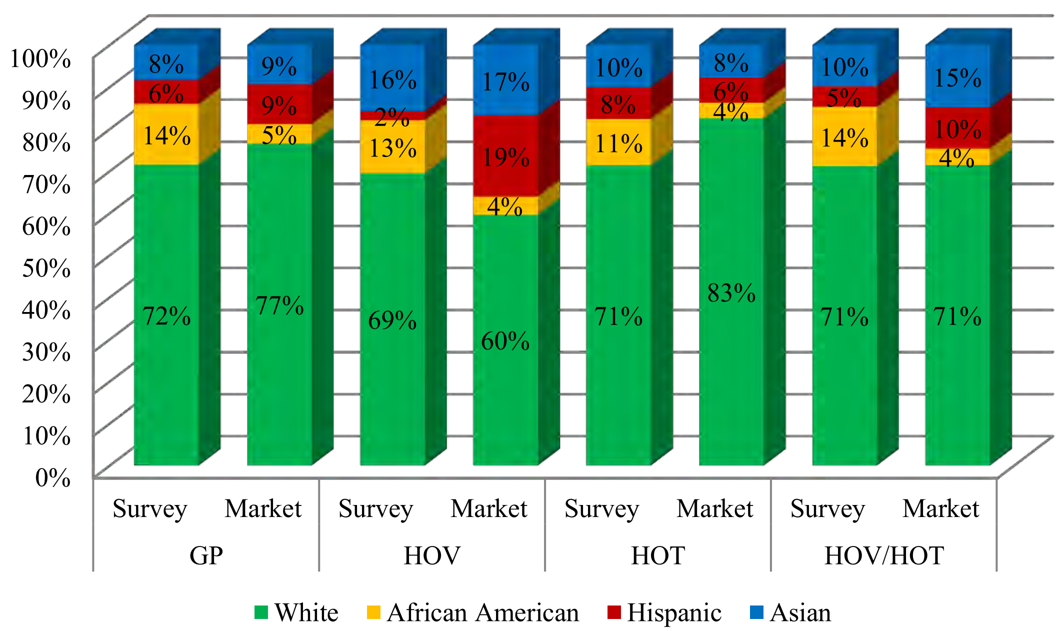


Figure 40 Households Ethnicity across Markets for Each Data Source

6.2.5. Demographic Profiles

Using marketing data, demographic profiles for household education, head of household age, gender, home ownership, dwelling type, and length of residence have been illustrated in the following figures. Because survey data are not available for these variables, no distribution based on survey data could be developed similar to previous graphs. Although the demographic profiles for income, vehicle ownership, household size, and ethnicity based on both marketing and survey data have been illustrated and explained earlier in this section, these profiles based only on marketing data have been re-illustrated here to generalize the presentation of the results. Figure 41 illustrated income profile across the markets; Figure 42 illustrates vehicle ownership profile across the markets; Figure 43 illustrated number of adults profile across the markets; Figure 44 illustrates number of children profile across the markets; and Figure 45 illustrates ethnicity profile across the markets.

Regarding education, the differences across the markets are also significant (Figure 46). A 21% increase in users with Bachelor or higher degrees in HOV/HOT market and 14% in HOT market compared to the GP and HOV markets was observed.

Regarding marital status, the differences across the markets are significant (Figure 47). HOV market represent 14% fewer married households and HOT market represents 9% more married households compared to the GP market. HOV/HOT market is not significantly different from GP market.

Regarding head of household age, the differences across the markets are significant (Figure 48). HOT and HOV/HOT markets include 12% more users with head of household range of 35-44 (mid-life age who may have the highest productivity) compared to GP and HOV markets. With respect to head of household gender, no significant and meaningful difference across the markets was observed (Figure 49).

Regarding home ownership, 95% or more of all the markets are owners. Considering the small portion of renters, HOT market has 40% fewer renters, and HOV/HOT market has 20% fewer renters (Figure 50). Similarly, 90% or more of all the markets live in single family dwelling units (Figure 51). Considering the small portion of multifamily dwelling units (including condos and apartments), HOV market represent 25% more households live in multifamily dwelling units and HOT market represent 25% fewer households live in multifamily dwelling units.

Households' length of residence profile is illustrated in Figure 52 across the markets. All the four markets represent almost equal percentages of households with less than six years of residence. HOV market represents 12% and HOV/HOT market represents 9% more households with six to ten years of residence. Accordingly they represent fewer numbers of households with ten years or more length of residence.

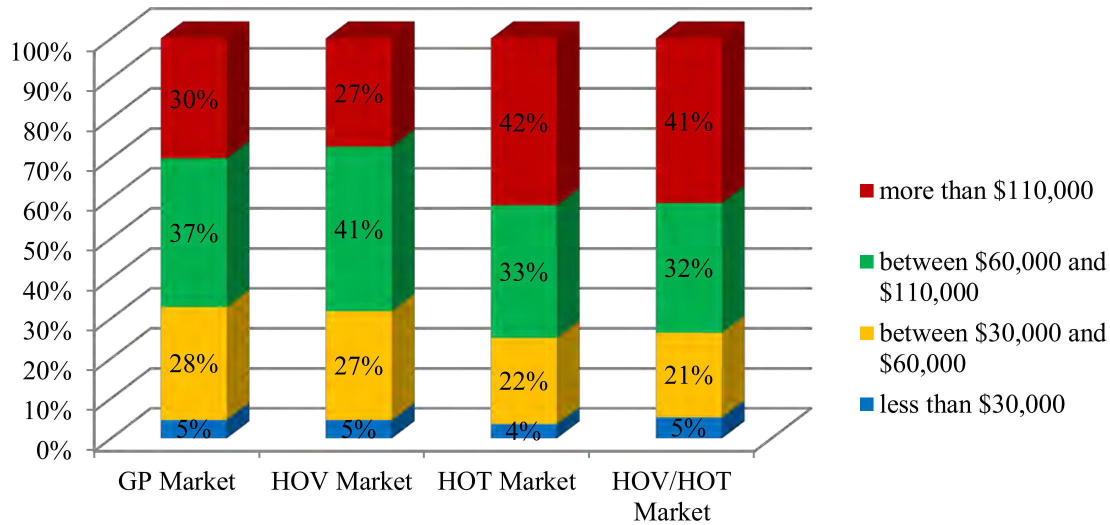


Figure 41 Households Income Profiles across the Markets (N=13,452)

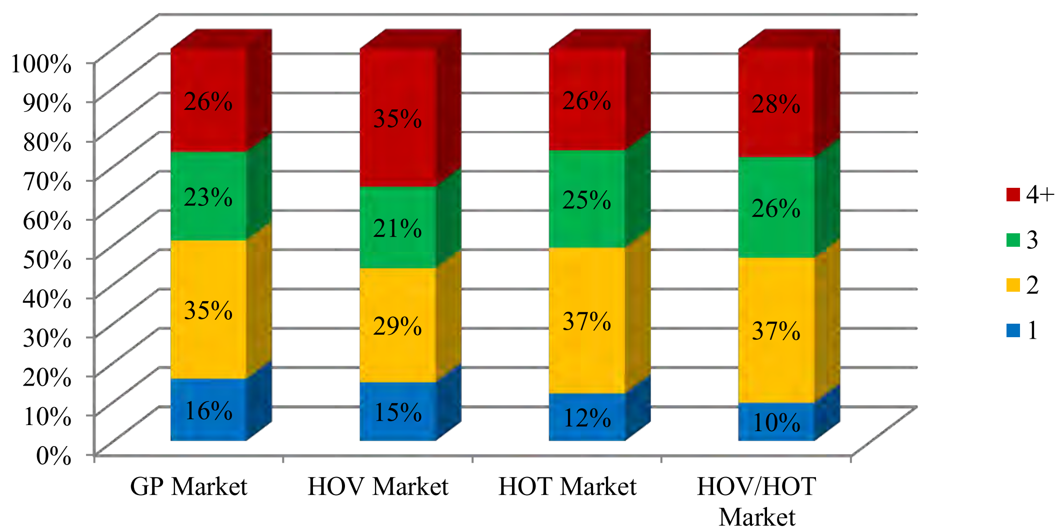


Figure 42 Households Vehicle Ownership Profiles across the Markets (N=13,477)

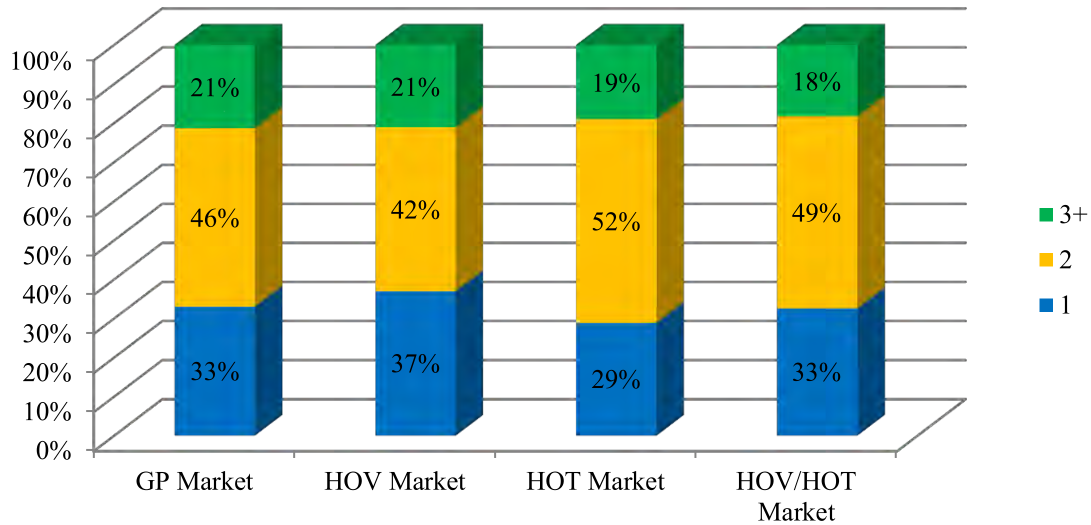


Figure 43 Households Number of Adults Profiles across the Markets (N=13,452)

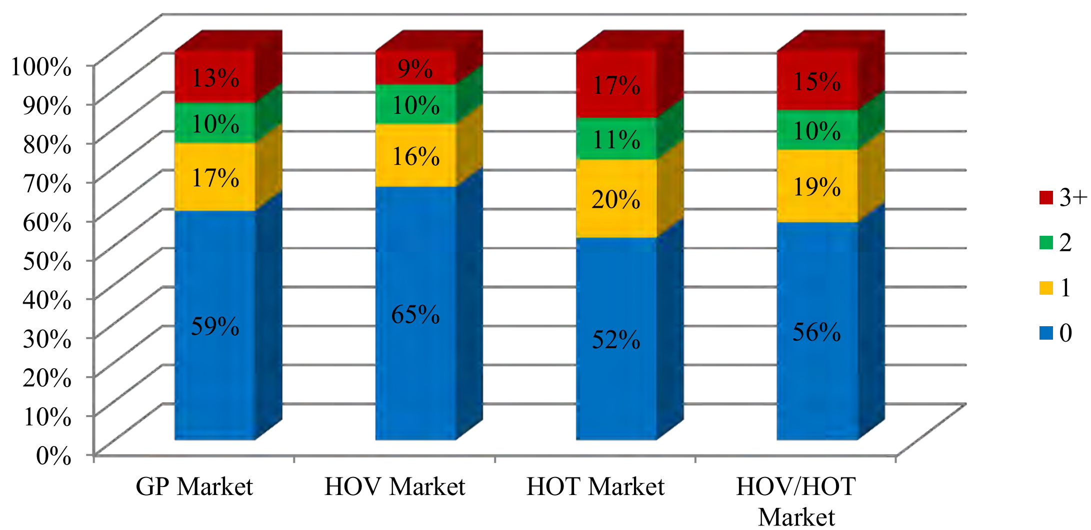


Figure 44 Households Number of Children Profiles across the Markets (N=11,097)

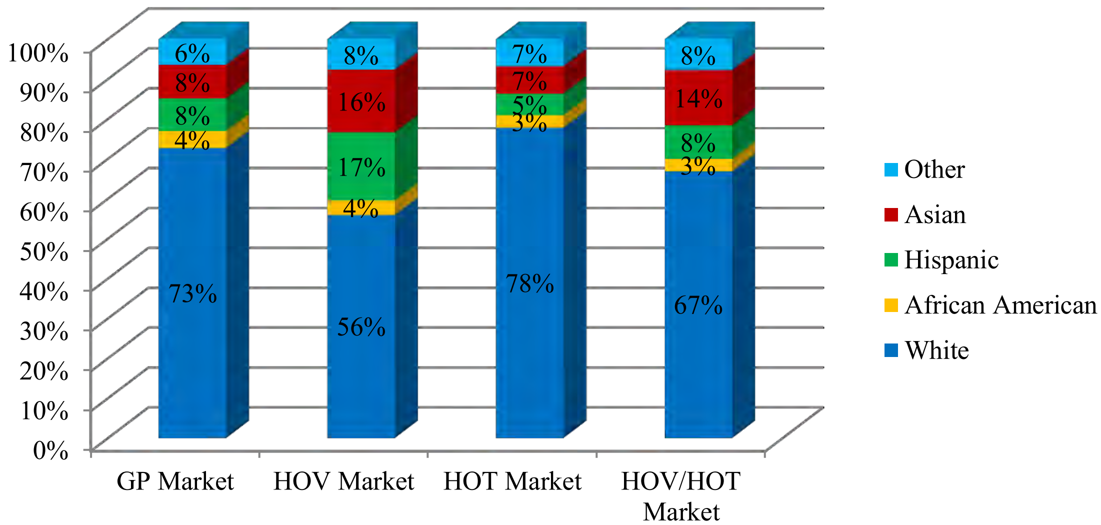


Figure 45 Households Ethnicity Profiles across the Markets (N=13,215)

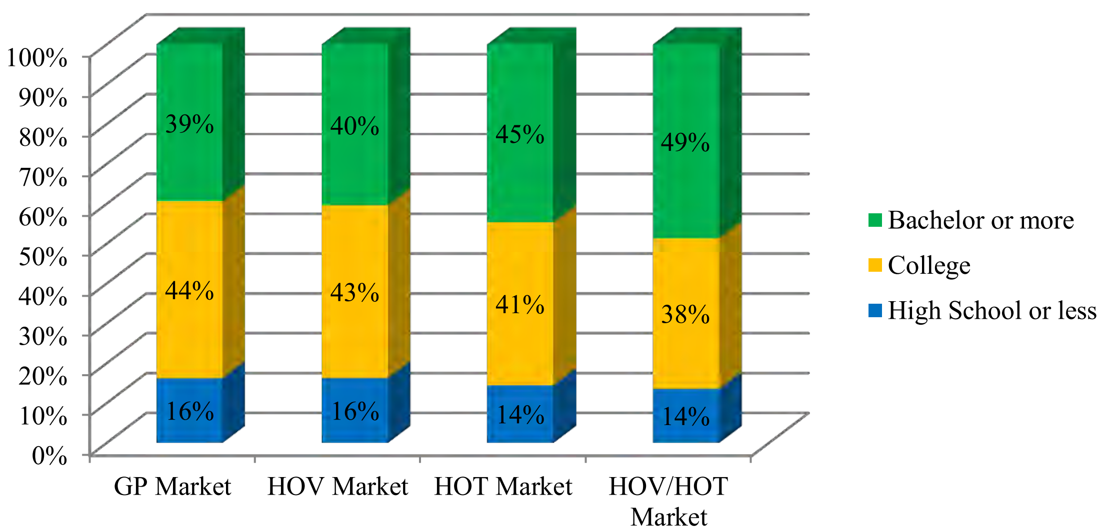


Figure 46 Households Education Profiles across the Markets (N=13,446)

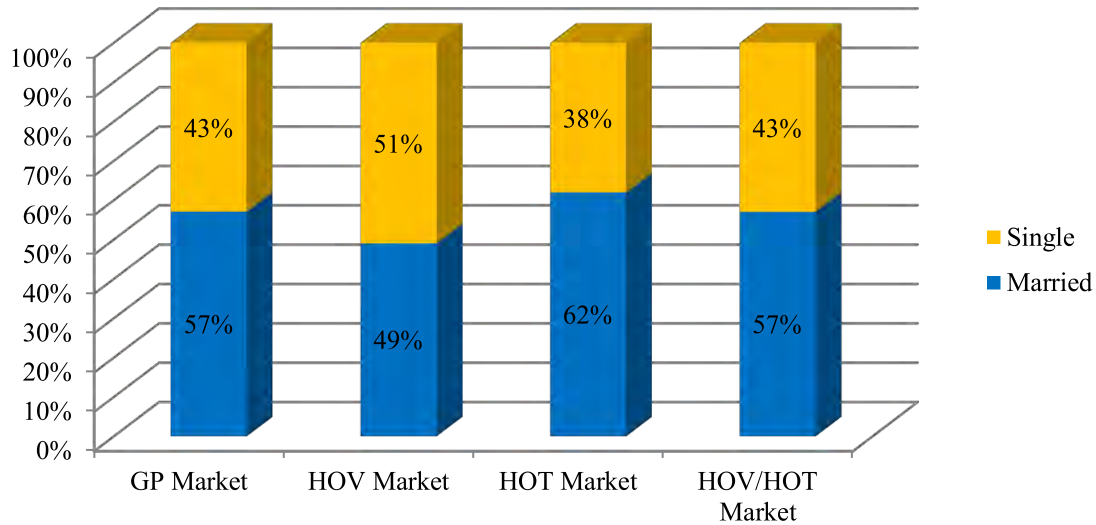


Figure 47 Households Marital Status Profiles across the Markets (N=13,144)

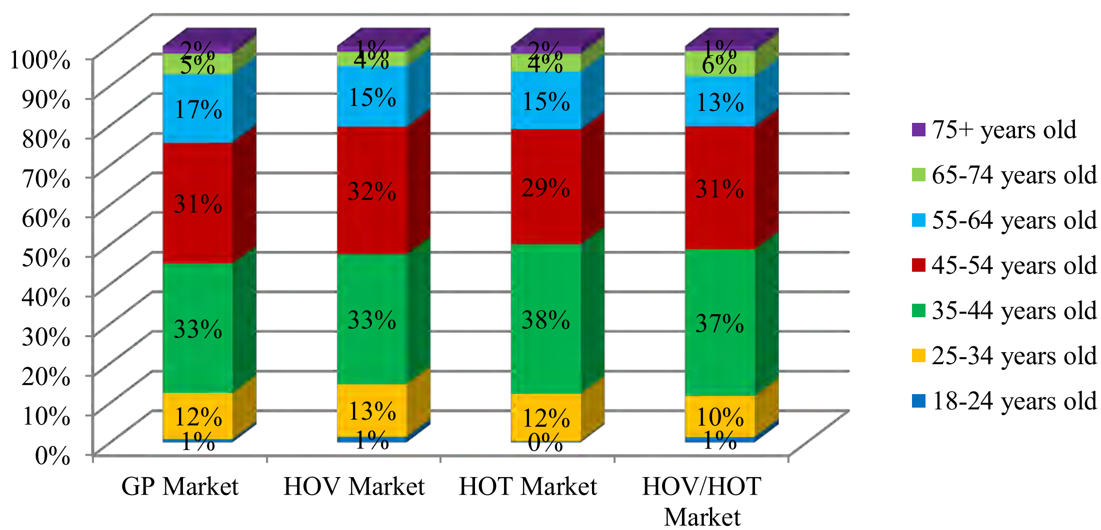


Figure 48 Head of Households Age Profiles across the Markets (N=13,452)

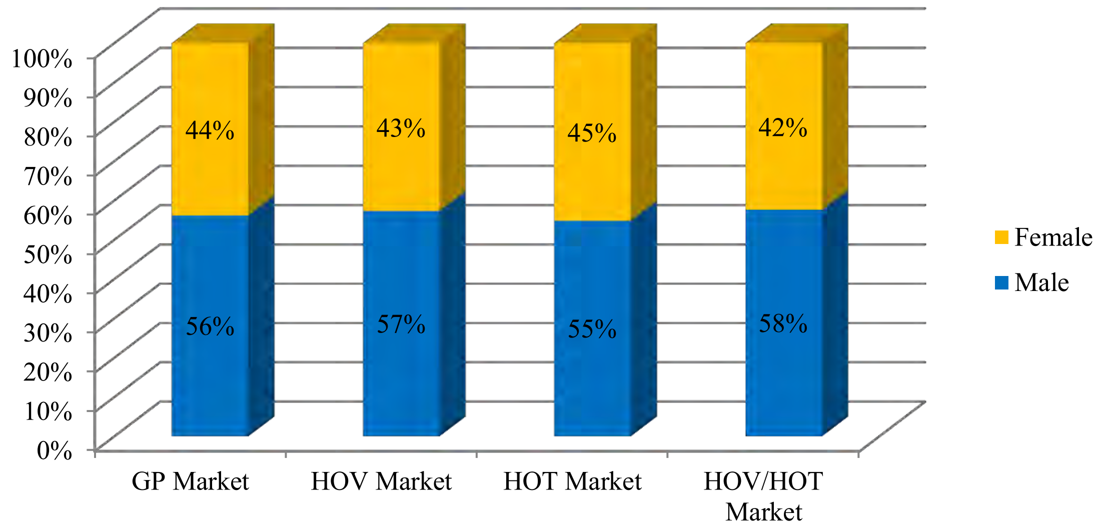


Figure 49 Head of Household Gender Profiles across the Markets (N=12,797)

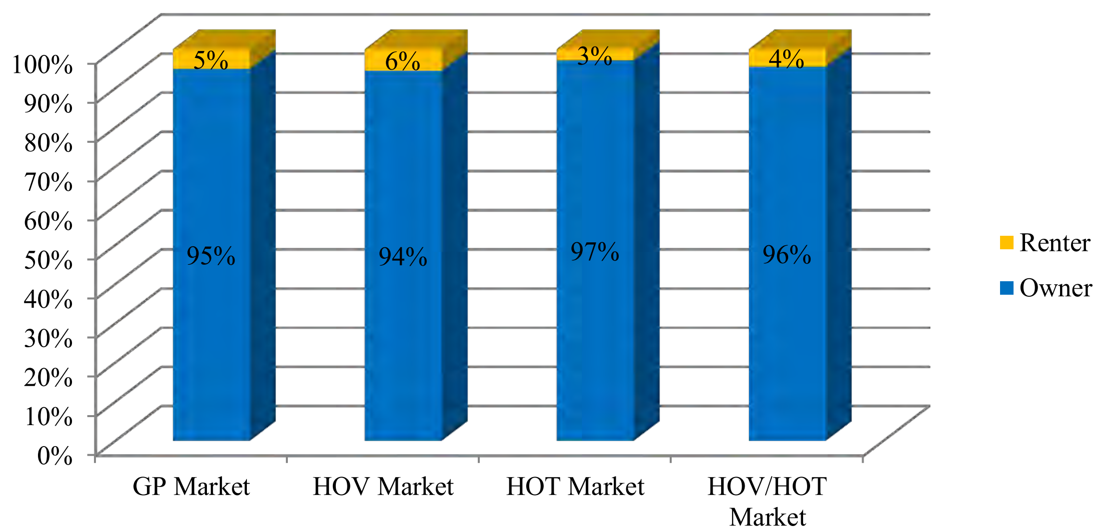


Figure 50 Households Home Ownership Profiles across the Markets (N=13,311)

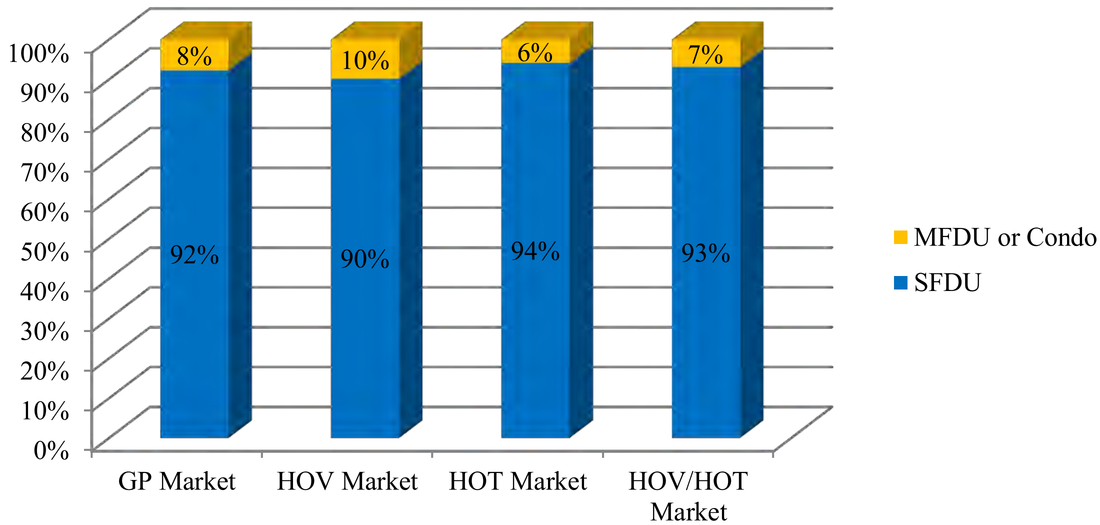


Figure 51 Households Dwelling Type Profiles across the Markets (N=13,452)

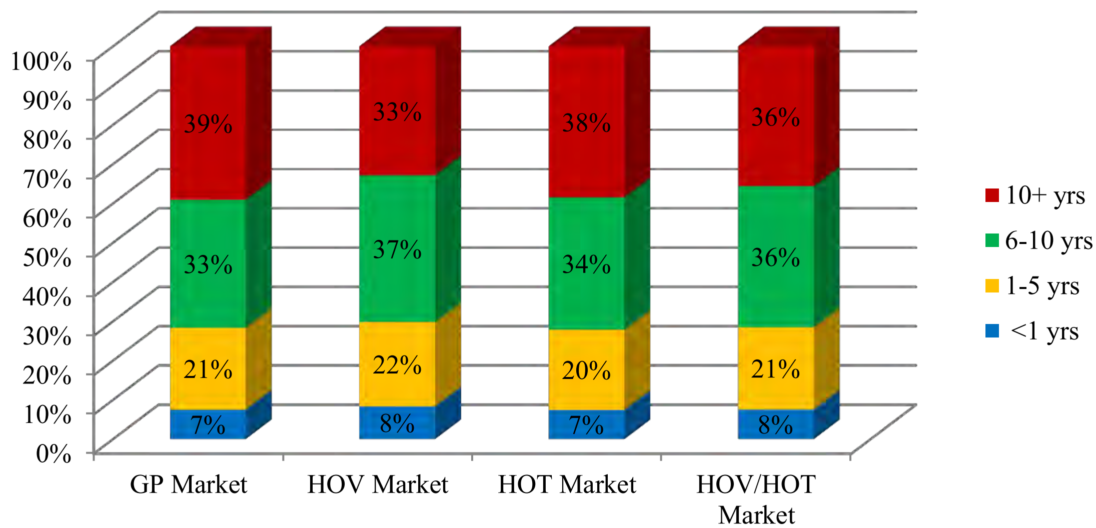


Figure 52 Households Length of Residence Profiles across the Markets (N=13,451)

6.3. Analysis of Variance

This section uses analysis of variance to assess which variables are appear to be significantly different across the established markets. Therefore, this analysis can by definition be applicable for only scale variables. However, ordinal variables and nominal variables, with only have two categories, such as marital status, home ownership, gender, and dwelling type have also been included. Hence, the only variable which has been excluded from this section is ethnicity.

Table 17 illustrated means and confidence intervals for each variable across the markets. The average and standard deviation estimates have already been presented and discussed for the primary variables (income, household size, and vehicle ownership). However, they are repeatedly presented here for the sake of comprehensiveness.

In terms of number of children HOT and HOV/HOT markets have the highest average number of children (1.09 and 1 respectively) implying that people in these markets have one children by average. On the other hand, HOV market has the lowest average number of children (0.69), which is 24% less than GP market.

Table 18 illustrates the statistical test results of equality of the means (ANOVA) across the markets, for all the variables in this study except for ethnicity. The mean values of all the variables are significantly different across the markets except for gender and adults. Across the significant variables, income has the largest F-value (implying the largest difference in mean across the markets), and respectively followed by marital status, number of children, vehicle ownership, education.

Table 17 Mean and Standard Deviation of Socioeconomic Variables across the Markets (using marketing data)

	GP Market (N=8,862)		HOV Market (N=1,189)		HOT Market (N=2,482)		HOV/HOT Market (N=919)	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Income	$71,030	$36,866	$67,162	$37,078	$81,263	$41,115	$77,801	$41,384
Vehicles	2.79	1.41	3.06	1.56	2.83	1.34	2.96	1.42
Adults	1.96	0.88	1.92	0.92	1.97	0.84	1.92	0.85
Children	0.91	1.45	0.69	1.17	1.09	1.53	1.00	1.51
Education	2.23	0.71	2.24	0.71	2.30	0.71	2.35	0.71
Marital Status (1=married, 2=single)	1.43	0.50	1.51	0.50	1.38	0.49	1.43	0.50
Age	3.75	1.16	3.63	1.13	3.67	1.13	3.68	1.14
Gender (1=male, 2=female)	1.44	0.50	1.43	0.50	1.45	0.50	1.42	0.50
Home Ownership (1=renter, 2=owner)	1.95	0.22	1.95	0.22	1.97	0.17	1.95	0.21
Dwelling Type (1=SFDU, 2=MFDU)	1.08	0.27	1.10	0.30	1.06	0.24	1.07	0.25
Length of Residence	7.08	1.89	7.22	1.95	7.71	1.91	7.73	2.09

Table 18 ANOVA: Test of Equality of Market Means

	F	Sig.
Income	63.186	0.000
Vehicles	15.114	0.000
Adults	1.374	**0.249**
Children	18.096	0.000
Education	12.924	0.000
Marital Status	18.129	0.000
Age	6.481	0.000
Gender	1.050	**0.369**
Home Ownership	5.270	0.001
Dwelling Type	6.623	0.000
Length of Residence	5.650	0.001

The changes in distribution as well as averages of all socioeconomic attributes of the frequent corridor commuters across the market have been illustrated visually and discussed. Meanwhile, there is a potential of inter-correlation between some of these variables. For example, high educated people are usually high income as well; hence, we cannot determine the higher percentage of high educated people in HOT market is the impact of income or it is the direct impact of the education using the presented descriptive statistics. To understand the real impact of each socioeconomic variable on users travel behavior independently, advanced statistical models will be developed at the end of this chapter and final conclusion will be derived.

6.4. Discussion about Application of Marketing Data

While the socioeconomic analytical results across the data sources are well correlated, some discrepancies have also been observed, particularly for household size which showed the lowest quality in the marketing data accuracy evaluation section.

While survey data presumed to report the true household demographics, the small size and potential sampling bias of the sample could significantly affect the results. For instance, the survey sample included only 100 households that fell into the HOV market, which appears to be the most sensitive user group of the study. Moreover, given the small sample size, it was not possible to statistically verify the differences across the user markets. Of course expanding sample size in future studies would potentially increase the accuracy and validity, but with higher survey costs.

Marketing data source could potentially be used in place of surveys, particularly when the trends of differences between groups are more the focus of attention rather than each individual household. Despite providing the large sample size at low cost, marketing data lack any travel behavior variable. Therefore, application of marketing data will likely be most useful when supplementing other travel information such as license plate surveys, RFID tag reads, or cell phone tracking data.

License plate data collection is becoming more convenient and less expensive using advanced technologies such as ALPRs (Automatic License Plates Readers) (Colberg, 2013). Whereas license plates can be matched to the registration location, there is no guarantee that all of the drivers actually live in the household where their vehicles are registered (Granell, 2002). Despite the advantages and disadvantages of application of license plate data, joining them to marketing data via vehicle registration database provided a very large across all the user market groups. Moreover, if errors in marketing data are random and normally distributed, a large enough sample based upon plate data should result in error cancellation and provide comprehensive and statistically significant demographic profiles across all the groups of users at substantially lower cost. The cost of license plate data collection and marketing data acquisition has been estimated as 20 cents per household which is significantly lower than households travel surveys ($200 per household for Atlanta 2011 household travel survey).

Lastly, although aggregate data are publicly available at no cost, the aggregation functions removed the desired level of variation across the cases since a single central value is assigned to all the households in a certain boundary. However, aggregated data could still show the trends among the population groups at a lower scale.

CHAPTER 7

HOUSEHOLD LEVEL MODELING

This chapter contains the main modeling process and results of the study at the household level based on marketing data. This chapter starts with explaining the dependent and independent variables. Then descriptive statistics and graphs show the impact of socioeconomic attributes on managed lane usage. Lastly generalized linear models (logistic regression and count models) will illustrate the underlying relationship between socioeconomic attributes and managed lane use travel behavior. The logistics regression investigates the assignments of frequent corridor users to the four established markets (GP market, HOV market, HOT market, and HOV/HOT market) based on socioeconomic attributes. Lastly, the count models investigates the impact of socioeconomic attributes on number of trips along HOV lane and HOT lane controlling for total number of trips along the corridor using all the observed households.

7.1. Dependent Variables

Generally, the dependent variables of this study are the indications of household travel behavior toward managed lanes. "HOV Usage" as the percent of the time that a household were observed along the HOV lane before the conversion, and "HOT Usage" as the percent of the time that a household were observed along the HOT lane after the conversion are calculated for each household and the relevant statistics are illustrated in Table 19. These two variables have been previously used to establish the four markets in Chapter 6. Additionally, the statistics of total observations along the corridor per household (Total Freq), total observations along the corridor, before the conversion, per household (Before Freq), total observations along the corridor, after the conversion, per household (After Freq), total observations along the HOV lane, per household (HOV Freq), total observations along the HOT lane, per household (HOT Freq) have been illustrated. Count models will be used in section 7.4 to model HOV and HOT lanes use frequencies as a function of total frequency and socioeconomic attributes.

Table 20 illustrates the same statistics only for the top 5% frequent households which have been observed in the corridor more than 20 times across all the lanes. Although they represent only 5% of the observed households, they account for 42% of the observed license plates. HOV and HOT Usage variables for frequent households have been used before in defining four user markets (GP market, HOV market, HOT market, and HOV/HOT market). Logistic regression modeling will be used in section 7.3 to illustrate the impact of socioeconomic attributes on belonging to each category.

Table 19 Dependent Variables Statistics (All households, N=241,466)

		HOV Usage	HOT Usage	Before Freq	After Freq	HOV Freq	HOT Freq	Total Freq
Mean		0.10	0.05	3.10	1.86	0.35	0.24	4.95
Mean Std. Error		0.001	0.000	0.014	0.008	0.003	0.003	0.020
Median		0	0	1	1	0	0	1
Mode		0	0	1	0	0	0	1
Std. Deviation		0.27	0.19	6.65	3.75	1.71	1.42	9.61
Variance		0.07	0.04	44.21	14.03	2.93	2.00	92.39
Range		1.00	1.00	168	78	96	43	196
Minimum		0.00	0.00	0	0	0	0	1
Maximum		1.00	1.00	168	78	96	43	197
Sum		NA	NA	747,977	448,456	84,951	56,864	1,196,433
Percentiles	5%	0	0	0	0	0	0	1
	10%	0	0	0	0	0	0	1
	25%	0	0	0	0	0	0	1
	50%	0	0	1	1	0	0	1
	75%	0	0	2	2	0	0	4
	90%	0.5	0	8	5	1	0	12
	95%	1	0.5	15	9	1	1	22

Table 20 Dependent Variables Statistics (Top 5% frequent, N=13,476)

		HOV Usage	HOT Usage	Before Freq	After Freq	HOV Freq	HOT Freq	Total Freq
Mean		0.10	0.16	24.68	12.90	2.54	2.37	37.58
Mean Std. Error		0.00	0.00	0.11	0.07	0.05	0.04	0.15
Median		0.00	0.00	22	12	0	0	32
Mode		0.00	0.00	16	9	0	0	21
Std. Deviation		0.20	0.28	13	8	6	5	17
Variance		0.04	0.08	178	62	37	24	298
Range		1.00	1.00	168	78	96	43	176
Minimum		0.00	0.00	0	0	0	0	21
Maximum		1.00	1.00	168	78	96	43	197
Sum		NA	NA	332,598	173,809	34,217	31,962	506,407
Percentiles	5%	0.00	0.00	10	2	0	0	21
	10%	0.00	0.00	12	4	0	0	22
	25%	0.00	0.00	16	8	0	0	25
	50%	0.00	0.00	22	12	0	0	32
	75%	0.07	0.22	30	17	2	2	44
	90%	0.38	0.66	41	23	8	9	60
	95%	0.62	0.83	48	27	14	14	71

7.1.1. Managed Lane Usage across the Socioeconomic Groups

The HOV lane and HOT lane usage 95 percentile mean confidence intervals are presented across socio-demographic groups in the following figures. Because the observations are not equal per households, weighted mean for HOV and HOT usage confidence intervals are calculated based on the observation frequency.

In terms of income, HOT usage substantially increases as income increase while HOV usage decreases slightly (Figure 53). This was expected considering the need to pay toll for HOT lane access for single and double occupant vehicles and the need to carpool for HOV lane access without toll requirement.

In terms of vehicle ownership, HOV lane usage increases as vehicle ownership increases while HOT lane usage does not show a similar linear relationship. Although

HOT usage increases from 1 to 2 vehicles per households, it decreases slightly from 2 to 3 and 3 to 4+ vehicles per household (Figure 54).

Because income is the dominating factor influencing vehicle ownership, the same graph has been developed within medium income group. The same pattern of differences has been observed within medium income group rejecting any potential correlation error. Because of the potential impact of household size on vehicle ownership the same graph has been developed for vehicle ownership per number of adults in the households (Figure 55). HOV usage doesn't change from lower than one to between one and two categories, while increases to more than two categories. HOT usage shows similar trends which is increasing in the beginning and decreasing at the end. Therefore, vehicle ownership will be treated as an ordinal variable (1, 2, 3, 4+) in the modeling to be able to differentiate between the impacts of each category of vehicle ownership.

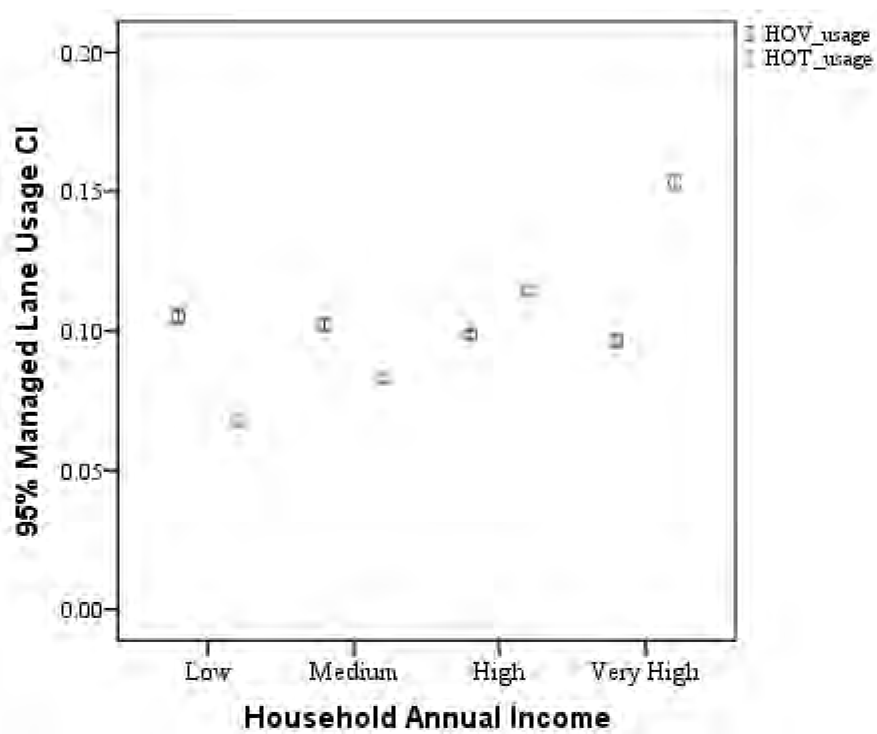


Figure 53 Managed Lane Usage Confidence Intervals across Income Groups (N=241,155)

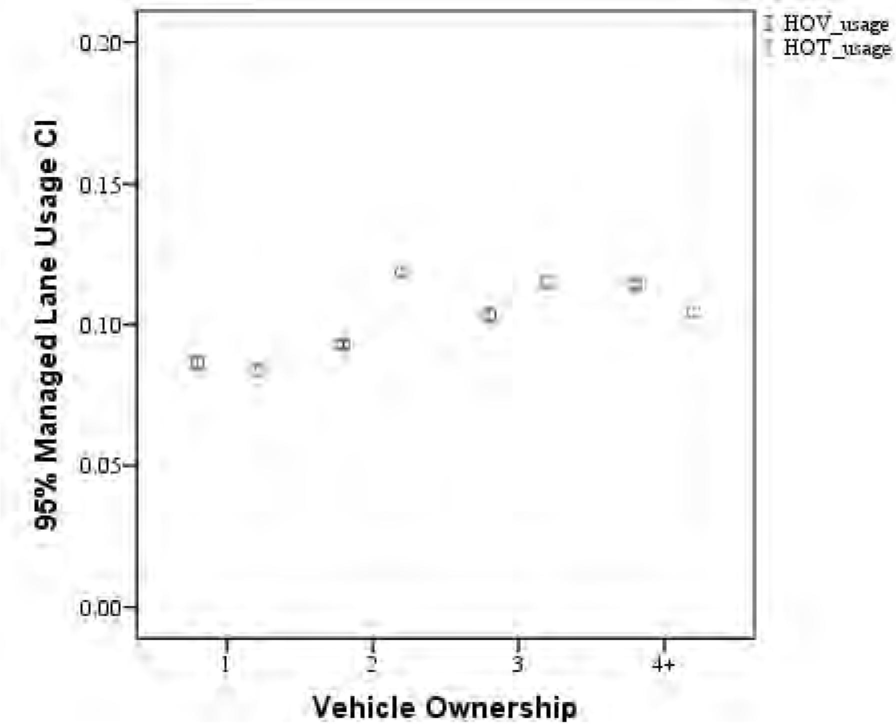


Figure 54 Managed Lane Usage Confidence Intervals across Vehicle Ownership (N=241,466)

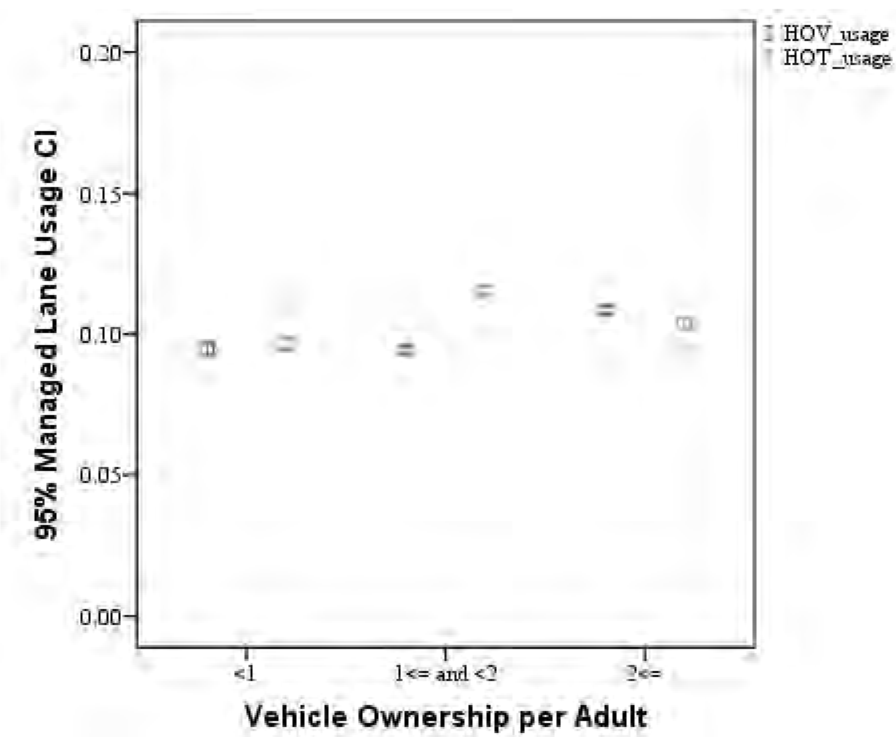


Figure 55 Managed Lane Usage Confidence Intervals across Vehicle Ownership per Adults (N=241,155)

In terms of household size, HOV usage does not practically changes across different number of adults and number of children (Figure 56; Figure 57). This is counter-intuitive to what we were expecting which is increasing carpooling as the household size increases. However, HOT usage increases as number of adults increase from one to two and decreases as number of adults increases to three and more. The third adult in the household might be grandparents who do not work and therefore, do not participate in commute travel.

Furthermore, HOT usage constantly increases intuitively, as number of children increases. Children may less likely participate in parents commute along the I-85 corridor. Kids usually go to schools closer to their house and do not commute with their parents to downtown. However, increase in number of kids is an indication of a busy life-style of parents with higher value of time to coordinate their responsibilities. Moreover, households with more number of kids are mainly in a financially established stage of the life cycle. Because of non-constant trends of difference number of adults and number of children will also be treated as ordinal variables in the modeling.

Impact of ethnicity is substantial on HOV and HOT usage. Hispanic, Asian, and African-American population HOT usage are significantly lower than White population (Figure 58). On the contrary, Hispanic and Asian HOV usage are significantly higher than White and African-American population.

HOT usage increases slightly as household level of education increases. However, HOV usage decreases slightly as household education level increase from high school to colledge and does not change beyond that (Figure 59).

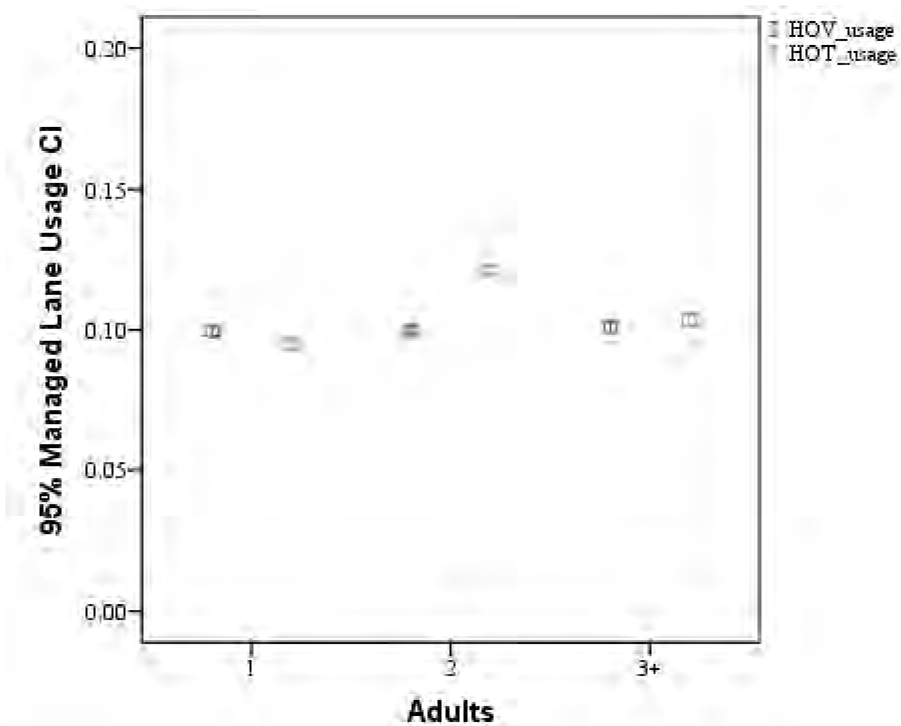


Figure 56 Managed Lane Usage Confidence Intervals across Number of Adults (N=241,155)

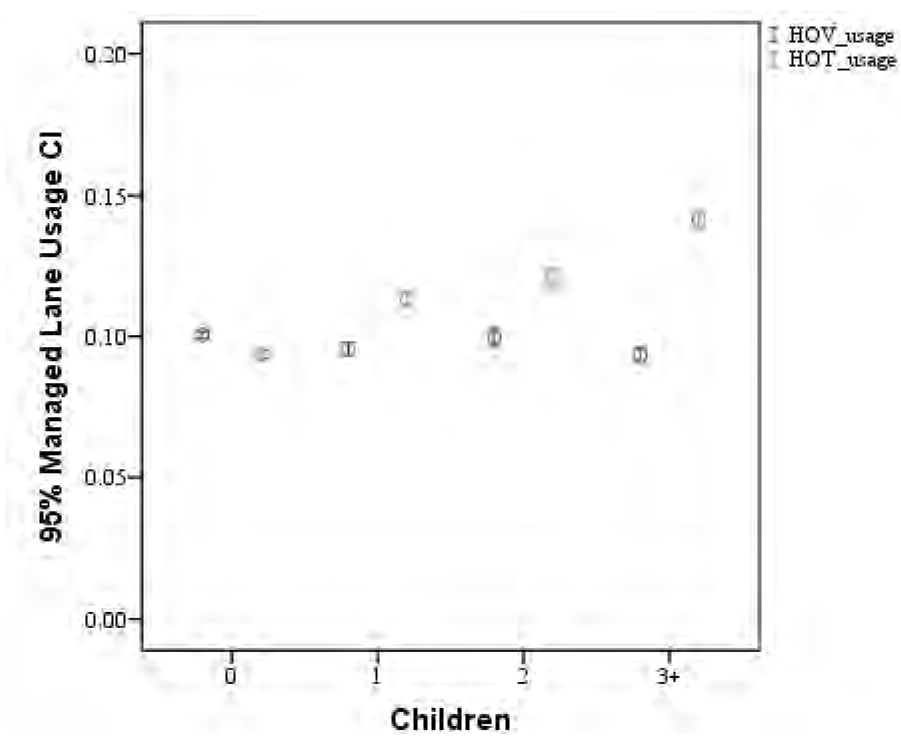


Figure 57 Managed Lane Usage Confidence Intervals across Number of Children (N=207,344)

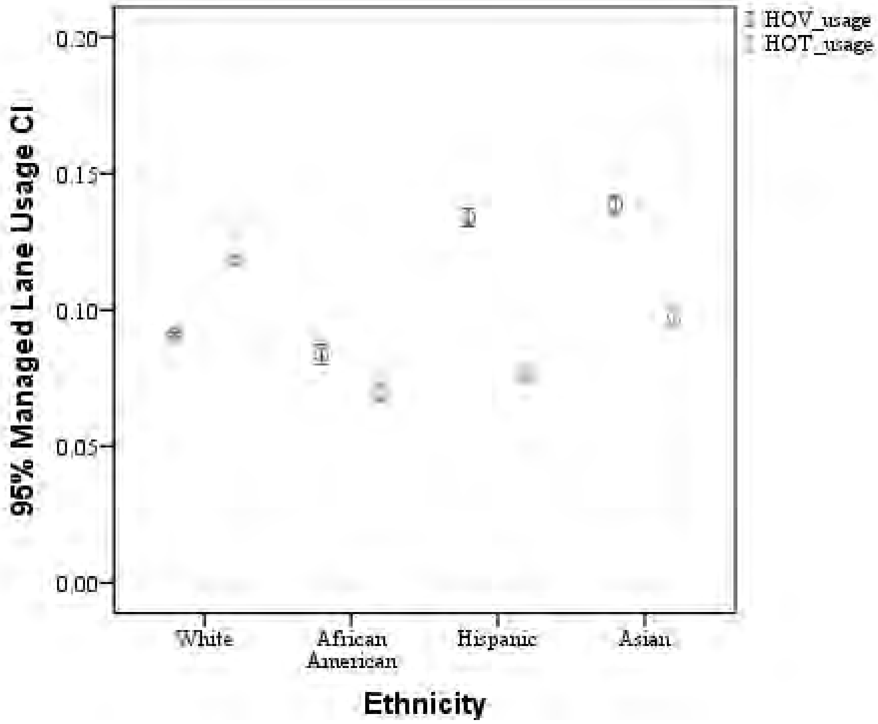


Figure 58 Managed Lane Usage Confidence Intervals across Ethnicity Groups (N=233,223)

HOT usage is significantly higher across married households compared to single households (Figure 60). This is intuitive considering higher income and larger households for married households. The HOV usage, however, is not significantly different between married and single couples. The impact of head of household age is not clear for both HOV and HOT usage (Figure 61). Therefore, the age groups have been aggregated to three groups in Figure 62. HOT usage is the highest at the middle age group and the lowest at two lower and higher age groups. This might be expected considering the higher value of time for people at most productivity ages (35-54 years old). HOV usage shows the similar trend with less significance. Head of household gender is not significantly impacting HOV and HOT usage (Figure 63). HOT usage is significantly higher across home owners compared to renters, which is expected considering the higher income of home owners (Figure 64). HOV usage is slightly higher across owners versus renters.

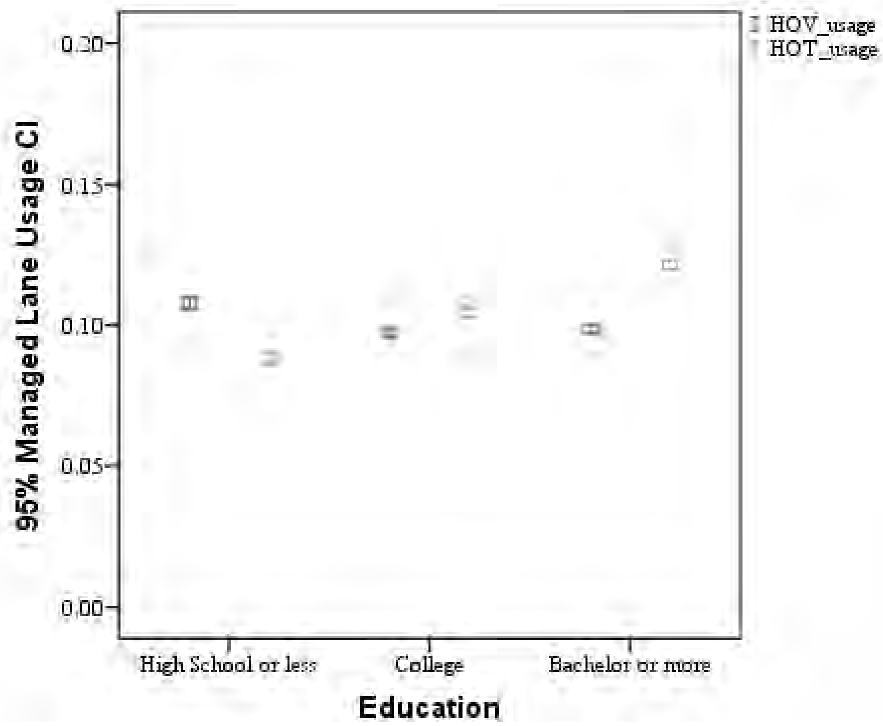


Figure 59 Managed Lane Usage Confidence Intervals across Education Groups (N=241,032)

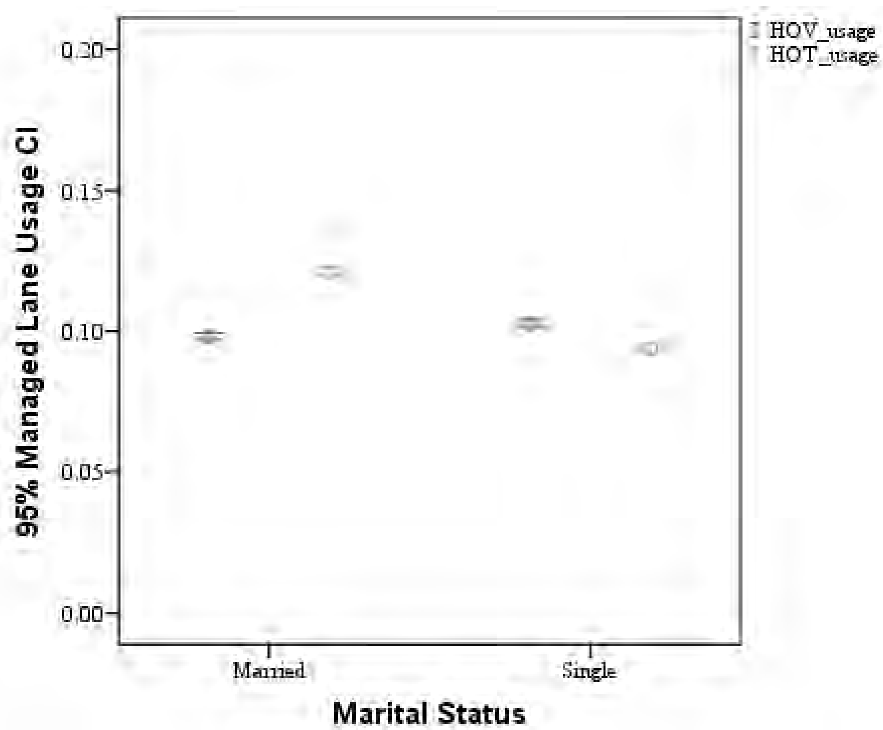


Figure 60 Managed Lane Usage Confidence Intervals across Marital Status (N=235,923)

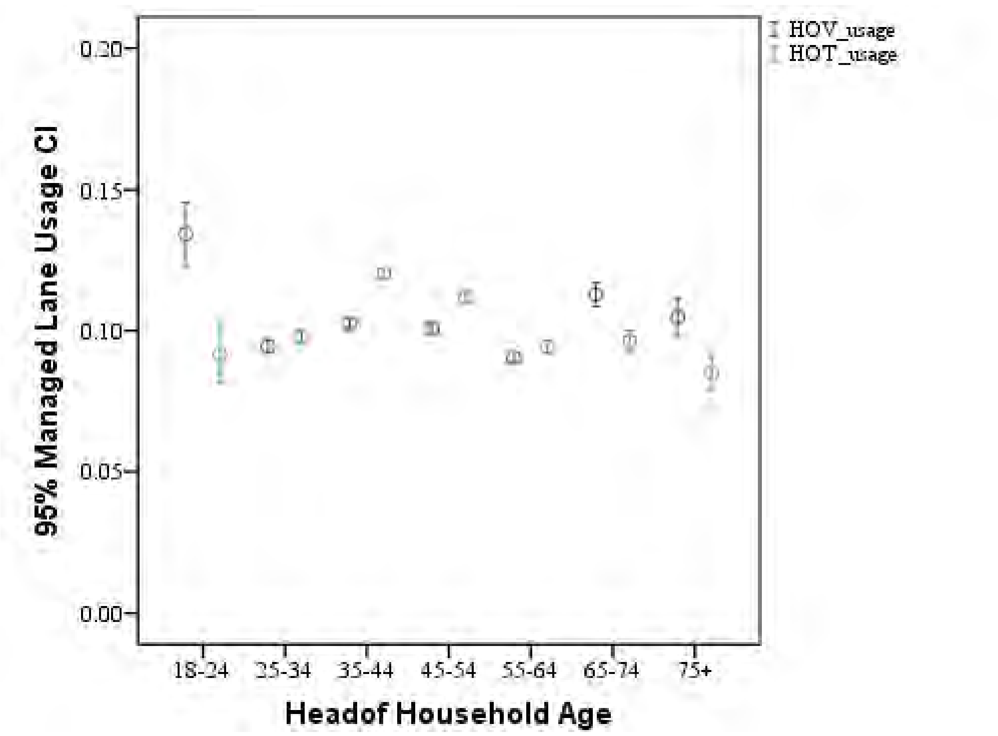


Figure 61 Managed Lane Usage Confidence Intervals across Head of Household Age (N=241,155)

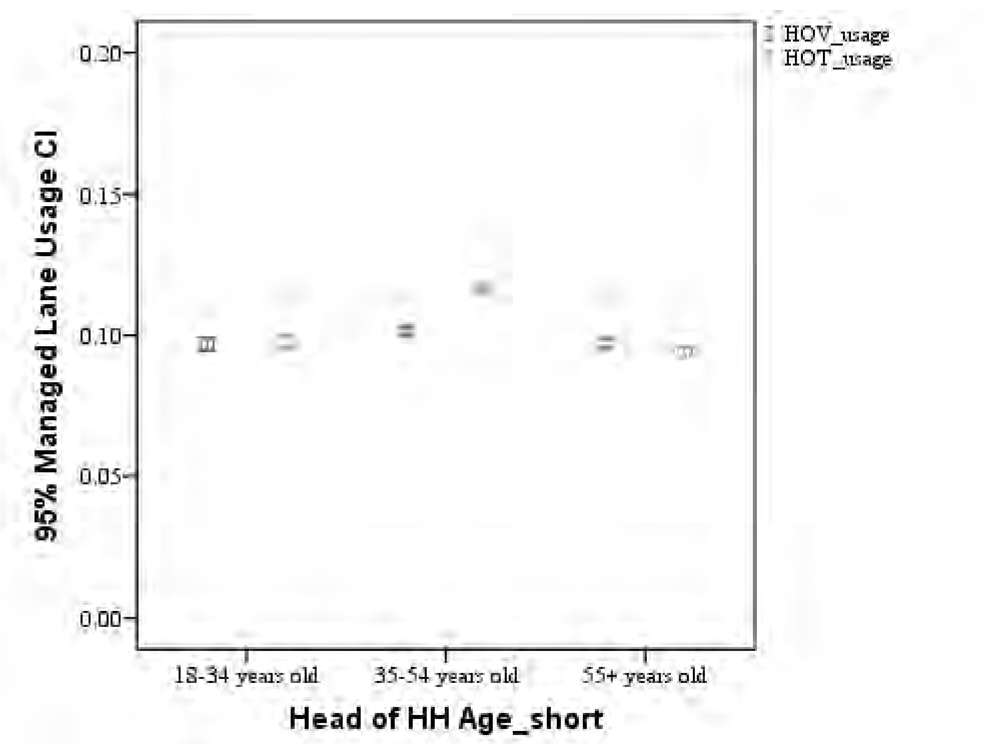


Figure 62 Managed Lane Usage Confidence Intervals across Head of Household Age (Short Version) (N=241,155)

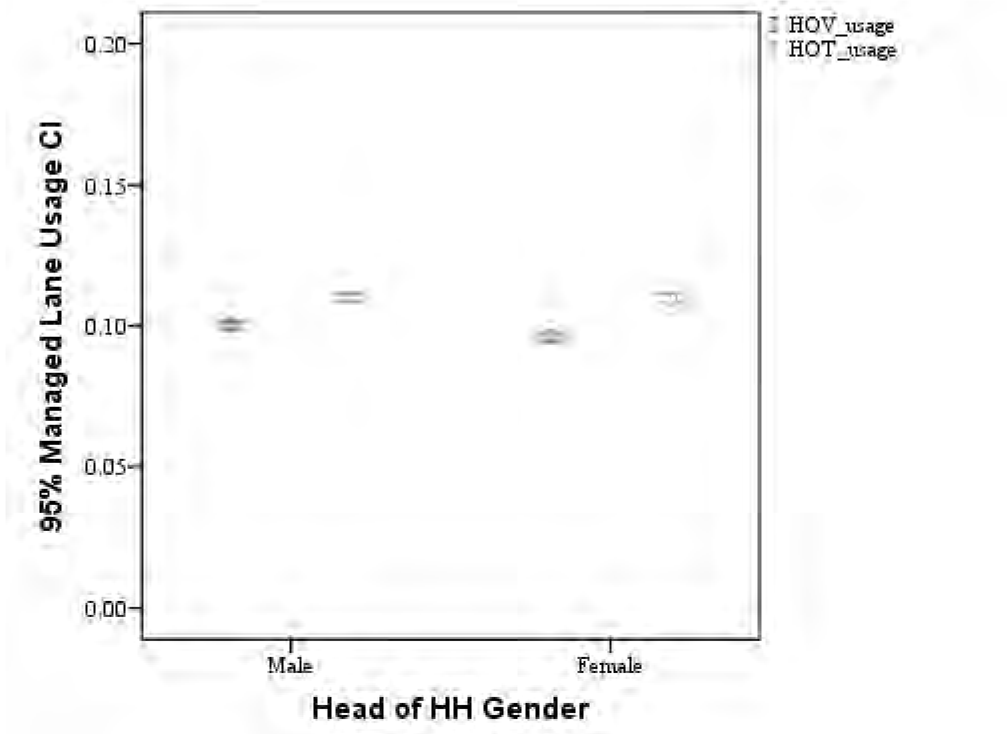


Figure 63 Managed Lane Usage Confidence Intervals across Head of Household Gender (N= 223,672)

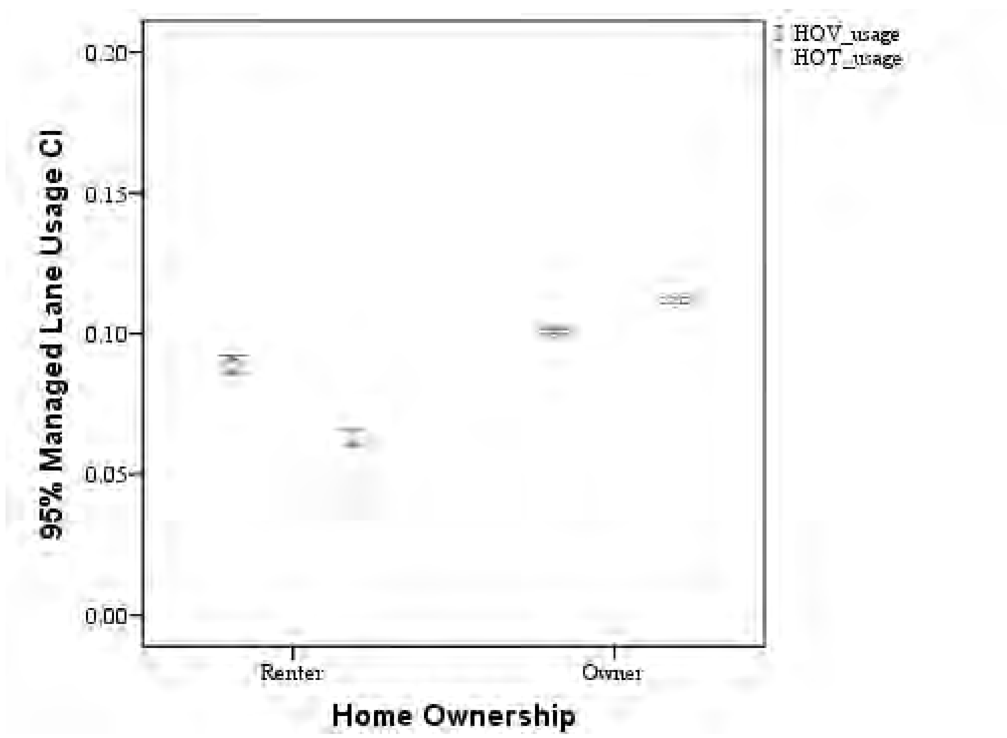


Figure 64 Managed Lane Usage Confidence Intervals across Home Ownership (N=238,211)

HOT usage is significantely higher at single family unit dwelling types compared to multi family unit dwelling types, condos and apartments (Figure 65). Similarly HOV usage is significantely higher at single family units with less amount of difference compared to HOT usage. This is intuitive considering the potential higher income and larger households living in single family units compared to multi family units, condos, and apartments.

HOV and HOT lanes usage increase slightly as length of residence increase up to 10 years and then slightly decreases (Figure 66).

Home age is substantially negatively correlated with HOT lane usage and similarly negatively correlated with HOV usage with less significance (Figure 67). This outcome is exactly similar to income and is intuitive. However, a similar trend of differences can be observed looking at home age within medium income group.

HOT usage is substantially increasing as household square footage increases, whereas HOV usage is not increasing with the similar trend (Figure 68). HOV usage does not increase significantely untill living area reaches 3000 square feet. Potential larger household size at very large houses increase the chance of carpooling and intuitively HOV usage.

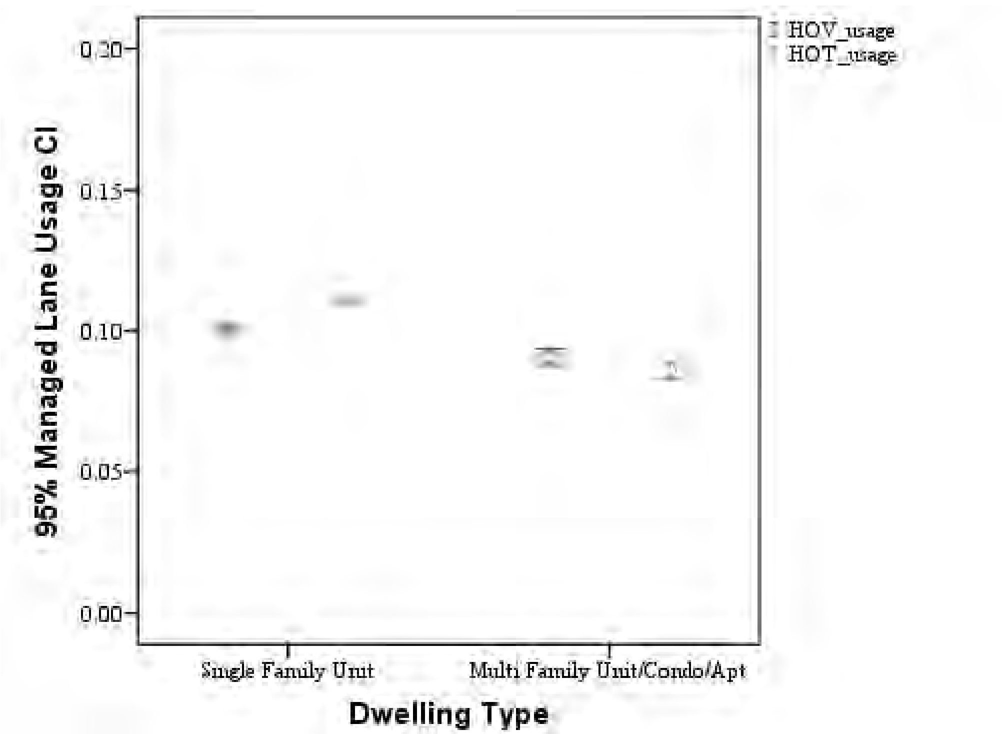


Figure 65 Managed Lane Usage Confidence Intervals across Dwelling Type (N=237,776)

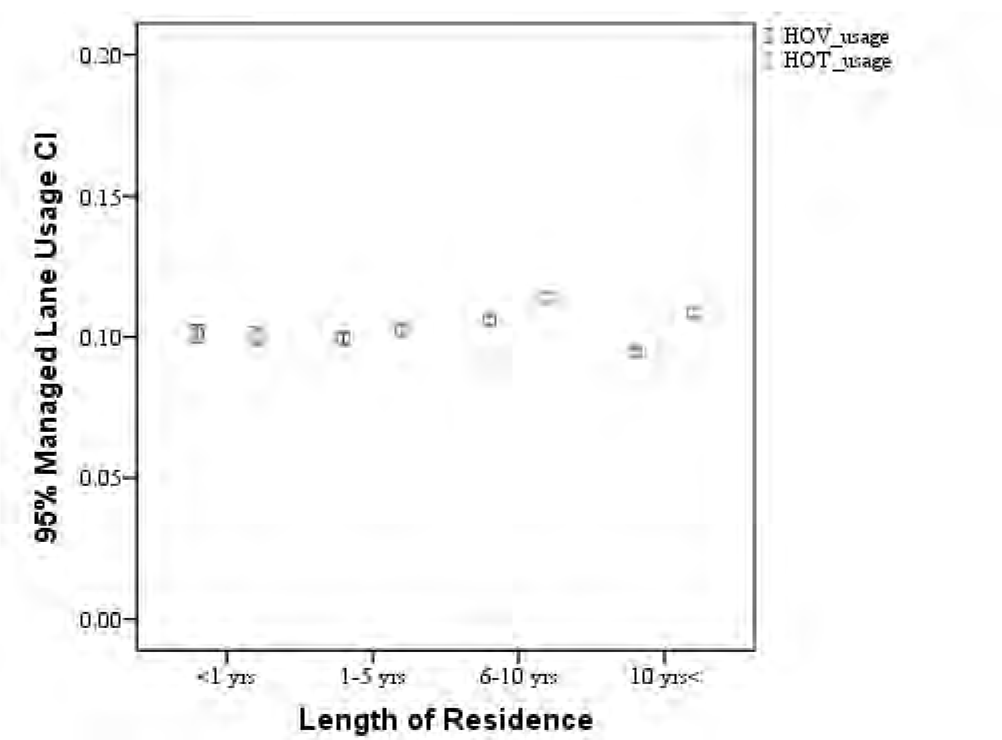


Figure 66 Managed Lane Usage Confidence Intervals across Length of Residence (N=541,155)

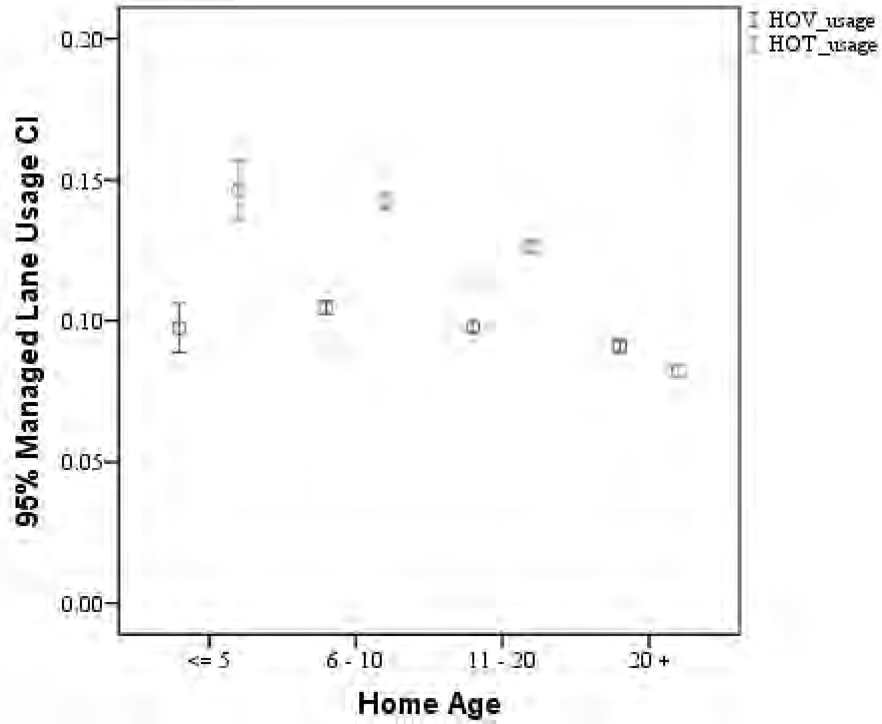


Figure 67 Managed Lane Usage Confidence Intervals across Home Age Groups (N=119,628)

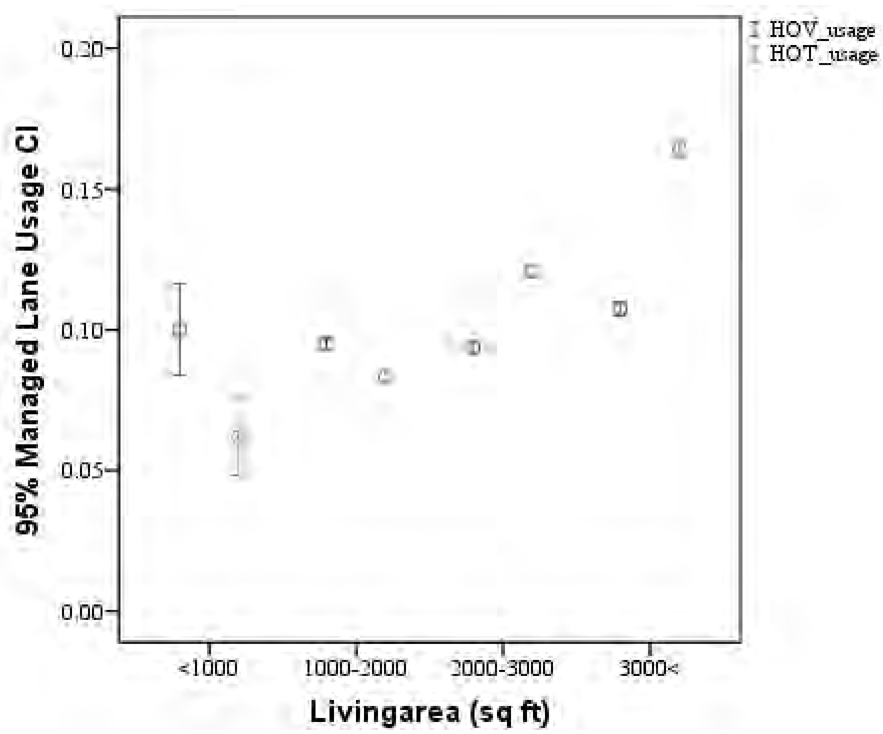


Figure 68 Managed Lane Usage Confidence Intervals across Living Area (N= 119,038)

All the explained managed lane usage confidence intervals across the socioeconomic groups are inter-dependent. For example, impact of marital status becomes insignificant if you look at all the households within a specific income group. Furthermore, some of the relationships might change if you look at only a partial of the population. Therefore, final conclusion, about the impact of socioeconomic attributes on managed lane usage, should only be based on statistical models which estimate the effect of each variable independent from other variables. Meanwhile, these descriptive statistics provides the preliminary information to better understand the underlying relationships.

7.2. Independent Variables

The complete list of variables that have been obtained from the marketing agency with detailed frequency tables are listed in the Appendix A. Table 21 illustrates the scale variables with minimum, maximum, mean and standard deviation estimations. The natural log transformations of income (divided by one thousand) have also been included. This transformation of the income is expected to show a better fit and more intuitive coefficients compared to the original income values; because it gives more sensitivity to the lower income population. More explanation will be provided later in this chapter, regarding income transformation.

Table 22 illustrates a list of ordinal variables with number and descriptions of the groups of each variable. Vehicle ownership and household size related variables have also been presented as ordinal variables because they will be used in this format in the developed models in this section. Head of household age have also been presented in short format for more convenient application and interpretation in the model. Table 23 illustrates list of nominal variables with number and descriptions of the groups of each variable.

Table 21 List of Scale Independent Variables

	Minimum	Maximum	Mean	Std. Deviation
HH Size	1	9	2.5	1.7
Adults	1	5	1.9	0.9
Children	0	9	0.7	1.3
Vehicle Ownership	1	7	2.8	1.5
Income	$10,000	$275,500	$65,810	$41,516
Income (Ln(1/1000))	2.30	5.62	3.9	0.6
Home Age	1	113	22	15.8

Table 22 List of Ordinal Independent Variables

	Groups	Description
Vehicle Ownership	4	1, 2, 3, 4+
Adults	3	1, 2, 3+
Children	4	1, 2, 3, 4+
Living Area	14	(1-749), (750-999), (1000-1249),…(6000-6999), (7000+) sq ft
Length of Residence	8	(0=6 mon), (7-12 mon), (1-2 yrs), …(16-20 yrs), (20+ yrs)
Head of HH Age	7	(18-24), (25-34), (35-44), (45-54), …, (65-74), (75+)
Head of HH Age	3	(18-34), (35-54), (55+)
Education	3	(High School or less), (Some college), (Bachelor or more)

Table 23 List of Nominal Independent Variables

	Groups	Description
Ethnicity	5	White, African American, Hispanic, Asian, Other
Head of HH Gender	2	Male, Female
Marital Status	2	Married, Single
Home Ownership	2	Owner, Renter
Dwelling Type	2	Single Family Unit, Multi Family Unit/Condo/Apt

7.2.1. Independent Variables Coverage

Not all the variables are available for all the 241,466 households in the study. Figure 69 illustrates the variables coverage. Almost all the variables have acceptable coverage except home age and living area which cover about half of the households. Although these two variables will be used in descriptive statistics of this section, they will not be inserted into the model because they decrease the models degree of freedom by about 50%. Furthermore, income is expected to cover most of their prediction power in the model.

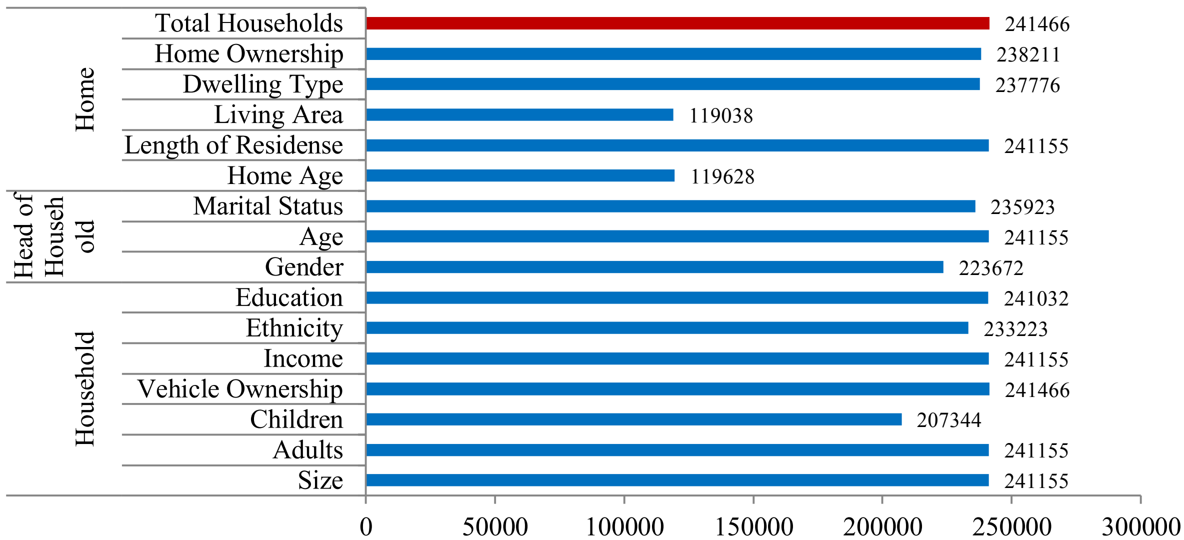


Figure 69 Independent variables coverage

7.2.2. Independent Variables Correlation

Diagnosing the potential correlations between the variables is important before developing the models. Table 24 illustrates the Pearson Correlations between the independent variables. Age of head of households, number of adults and length of residence are all inter-correlated with Pearson correlation around 0.3 and 0.4. Number of adults is also correlated with income with Pearson correlation of about 0.3. Household education and living area is correlated with income with Pearson correlation of about 0.4. The mentioned correlations are all less than 0.5 and do not cause the elimination of these variables from the model. However, this amount of correlation may have impact on the model coefficients which should be evaluated. Furthermore, all forms of income are highly correlated (more than 0.9). In most of the models only one form of income will be used. However, if there is any nonlinear relationship between income and dependent variables (Kitamura et al., 1997), two forms of income variables need to be maintained in the model.

Table 24 Independent Variables Correlation Matrix

	Vehicle Own	Age	Adults	Children	Home Age	Length of Res	Living Area	Education	Income	Income_sqrt	Income_Ln
Vehicle Own	1	.11	.21	.05	.04	.17	.10	-.08	.10	.10	0.10
Age		1	.30	-.10	.20	.43	.03	.01	.18	.19	0.18
Adults			1	.28	.05	.42	.14	.01	.30	.31	0.29
Children				1	-.11	.19	.14	.00	.23	.24	0.23
Home Age					1	.25	-.32	.02	-.08	-.11	-0.14
Length of Res						1	-.05	.01	.28	.30	0.30
Living Area							1	.28	.47	.48	0.46
Education								1	.37	.39	0.40
Income									1	.98	0.91
Income_sqrt										1	0.97
Income_Ln											1

7.2.3. Demographic Profiles across the Lanes

This section illustrates and explains the demographic profiles of users across the lanes using all the households. Table 25 illustrates mean and standard errors of the mean for the scale variables across the lanes. Considering the large sample size, the standard errors are substantially low. HOT lane average income ($77,956) is $10,545 higher than the adjacent GP lanes ($69,911), which represents 15% difference, and $10,152 higher than the HOV lane ($67,804), which also represents 15% difference.

Number of adults is only very slightly lower for GP lanes' users after the conversion. Average number of children along the HOT lane is very close to one (0.98) and is 27% higher in HOT lane compared to adjacent GP lanes and HOV lane. HOV lane has the highest vehicle ownership which accounts for 5% difference compared to adjacent GP lanes. However, HOT lane vehicle ownership is not different from adjacent GP lanes.

Table 25 Mean and Mean Standard Error of the Scale Variables across the Lanes

	HOV	HOT	GP Before	GP After
Income	$67,804 ($137)	$77,956 ($181)	$69,911 ($48)	$67,711 ($62)
Adults	1.91 (0.003)	1.91 (0.004)	1.91 (0.001)	1.88 (0.001)
Children	0.77 (0.005)	0.98 (0.007)	0.83 (0.002)	0.78 (0.002)
Vehicles	2.94 (0.005)	2.82 (0.005)	2.79 (0.002)	2.82 (0.002)

A household annual income profile across the lanes is illustrated in Figure 70. The income is presented in four groups: low income (less than $30,000), medium income (between $30,000 and $75,000), high income (between $75,000 and $120,000), and very high income (more than $120,000). General purpose lanes (before and after the conversion) and HOV lanes represent very similar income profiles; however, HOT lane represent a different income profile with about 10% of users who moved from low/medium income groups to very high income group across the HOT lane. In other words, the HOT lane has about 40% more very high income users; almost the same number of high income users, 22% fewer medium income users, and 28% fewer low income users.

Household vehicle ownership profile across the lanes is illustrated in Figure 71. Vehicle ownership distributions are quite similar across the lanes with minor differences. HOV lane represents 15% more households with four and more vehicles and 8% fewer households with two vehicles compared to the adjacent general purpose lanes. HOT lane represents 22% fewer households with one vehicle ownership and 9% more households with two vehicle ownership compared to the adjacent general purpose lanes.

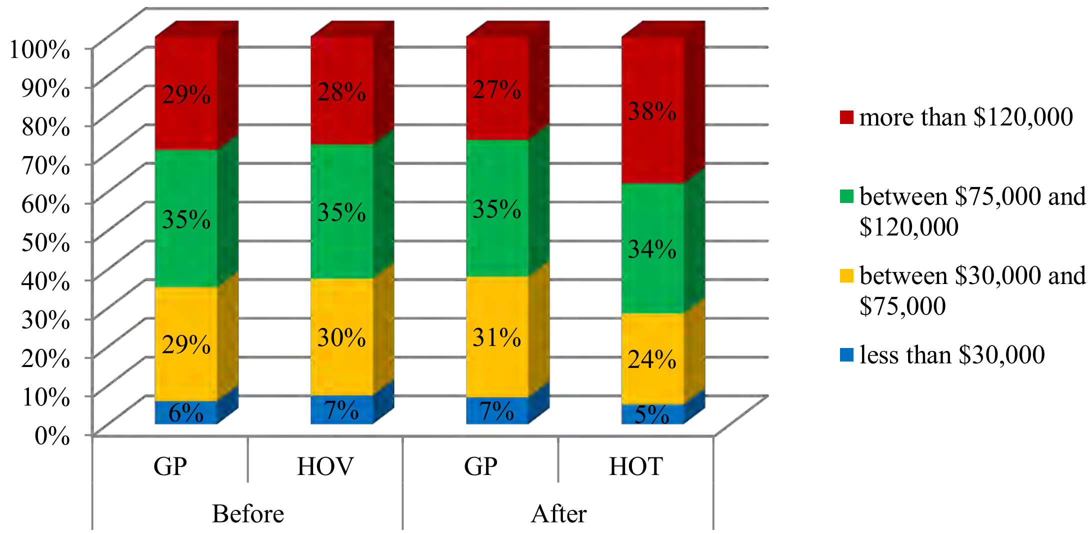


Figure 70 Households Annual Income Profiles across the Lanes (N= 241,155)

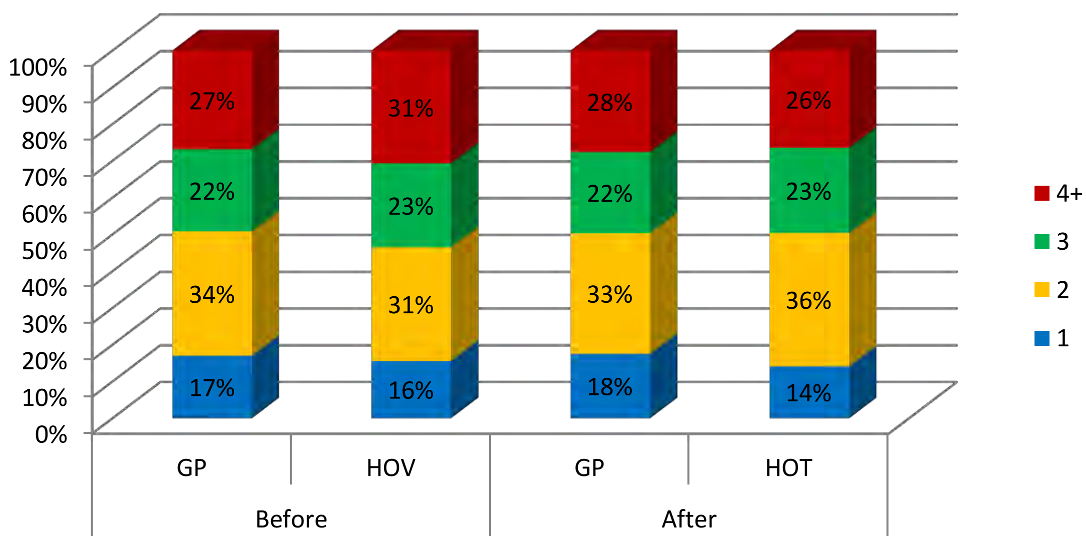


Figure 71 Households Vehicle Ownership Profiles across the Lanes (N= 241,466)

Households' number of adults profile across the lanes is illustrated in Figure 72. HOT lane represents a different distribution by 13% fewer households with one adult and 11% more households with two adults compared to the adjacent general purpose lanes. HOV lane is very similar to the adjacent general purpose lanes in terms of number of adults.

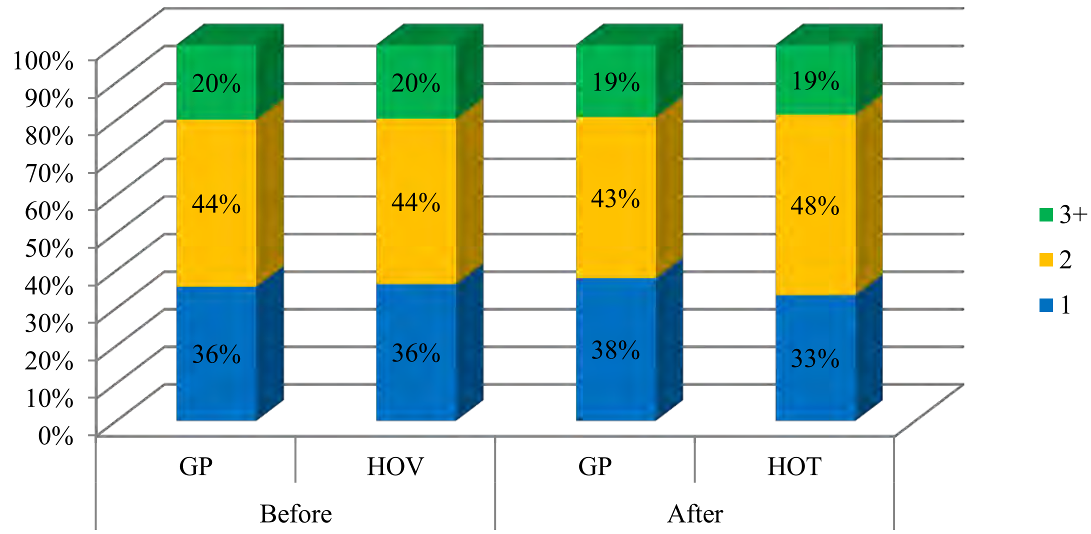


Figure 72 Households Number of Adults Profiles across the Lanes (N= 241,155)

Households' number of children profile across the lanes is illustrated in Figure 73. HOT lane represents a different distribution by 11% fewer households with no children and about 12% more households with one or two children and 36% households with three or more children compared to the adjacent general purpose lanes.

Although this outcome was expected considering the higher chance of carpooling in addition to higher income in larger households especially in those with children, the fact that HOV households' number of adults and number of children distributions are not very different from general purpose lanes is unexpected. This may imply that the observed changes in HOT lane users household size is more the outcome of income. The final conclusion in this regard will be based on the statistical modeling which investigates the impact of each variable independent from other variables.

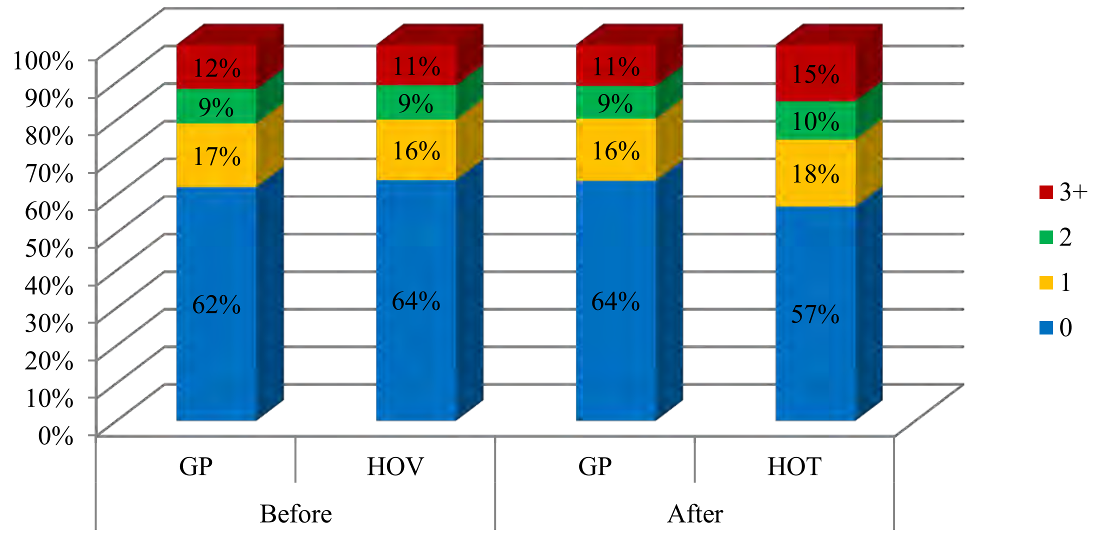


Figure 73 Household Number of Children across the Lanes (N= 241,155)

Households' ethnicity profiles across the lanes are illustrated in Figure 74. The HOV lane represent 50% more Asian and 33% more Hispanic population, and accordingly 8% fewer White population compared to the adjacent general purpose lanes. On the other hand, HOT lane represents 8% more White population, and accordingly 28% fewer African-America, 33% fewer Hispanic, and 12% fewer Asian population.

Therefore, the response to the HOV to HOT conversion appears to either have been different across these ethnicities, or across other socioeconomic, employment, spatial, lifestyle, or other variables that highly correlated with these ethnicities. Although there are different socioeconomic variables in the model, not all of the potential lifestyle and travel behavior variables are available and therefore their impacts may be lumped in with ethnicity variables.

Households' education profiles are illustrated in Figure 75. HOV lane represents 16% more households with high school or lower education level, whereas HOT lane represents 14% fewer households the same education level compared to the adjacent general purpose lanes. Furthermore, HOT lane represents 13% more households with Bachelor or more education level. This was expected, considering the fact that value of time for high educated people is generally higher.

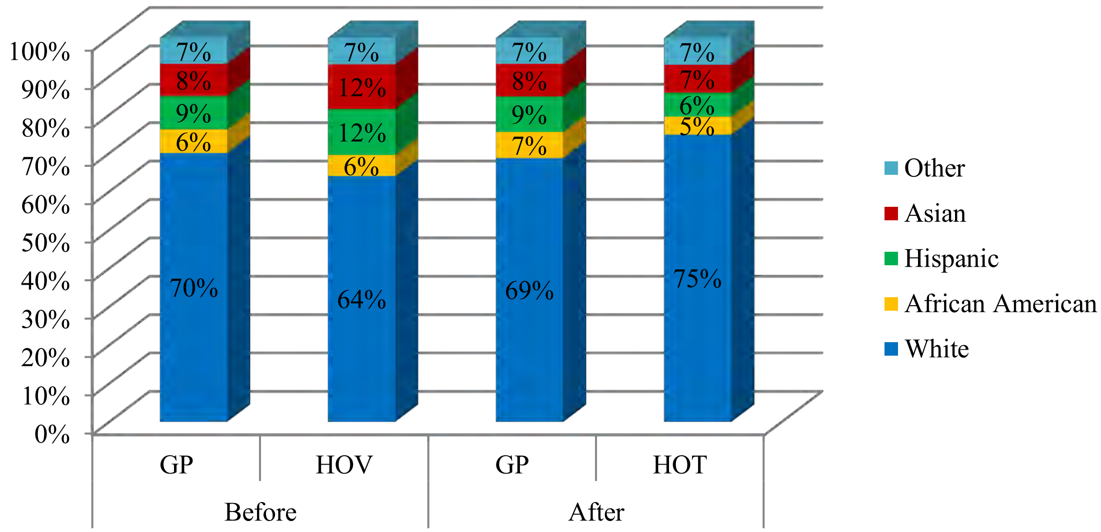


Figure 74 Households Ethnicity Profiles across the Lanes (N= 233,223)

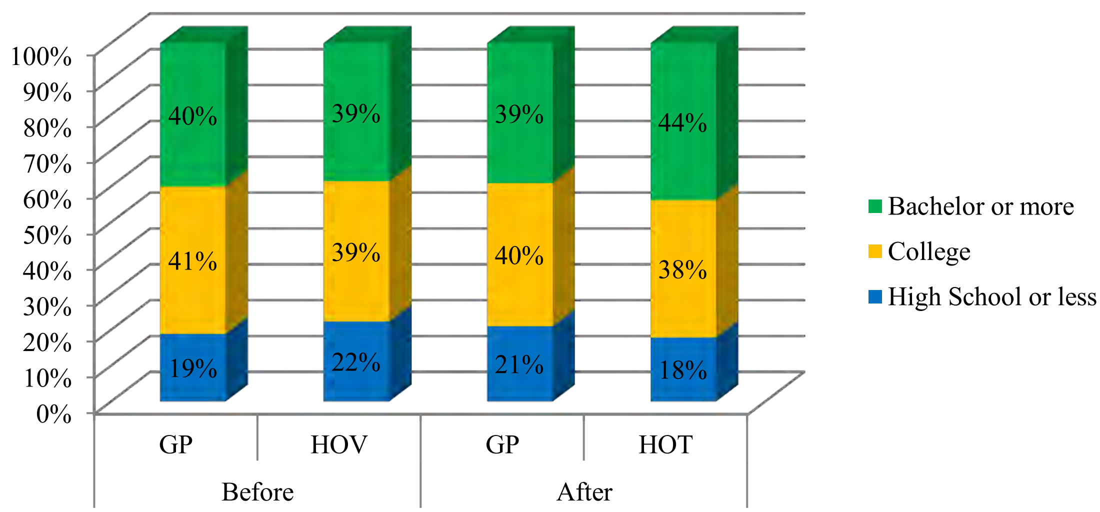


Figure 75 Households Education Profiles across the Lanes (N= 241,032)

Households' marital status across the lanes is illustrated in Figure 76. Whereas HOV lane is very similar to adjacent general purpose lanes illustrating close to 50-50 relationship, HOT lane represent 14% more married households. This outcome is expected, because married households are usually larger in terms of household size and have higher combined household income.

Head of household age is illustrated in Figure 77. Head of household age is not substantially changing across the lane. The only slight difference is that HOT lane has 9% more users with head of household age of 35-44 years old and accordingly 12% fewer users with head of household age of 55-64 years old. Head of household gender profile across the lanes is illustrated in Figure 78 and no meaningful difference could be observed. Households' home ownership profile across the lanes is illustrated in Figure 79. More than 90% of the corridor users are home owners (based on marketing data). Considering a small portion of renters (9%), HOT lane has 44% fewer renters compared to the adjacent general purpose lanes, which likely correlates to the impact of income.

Households' dwelling types across all the lanes are illustrated in Figure 80. Considering the urban structure of north-east Atlanta, it is not surprising that more than 90% of dwelling types are single family units. No significant change could be observed comparing HOV and HOT lanes to adjacent GP lanes regarding dwelling type.

Lastly, households' length of residence profile across the lanes is illustrated in Figure 81. No significant change can be observed across the lanes in terms of length of residence.

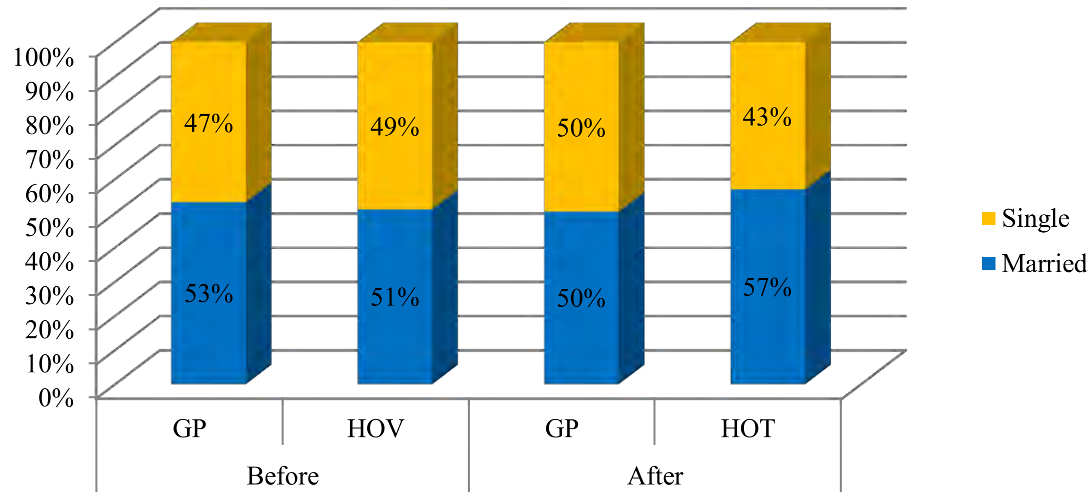


Figure 76 Households Marital Status Profiles across the Lanes (N= 235,923)

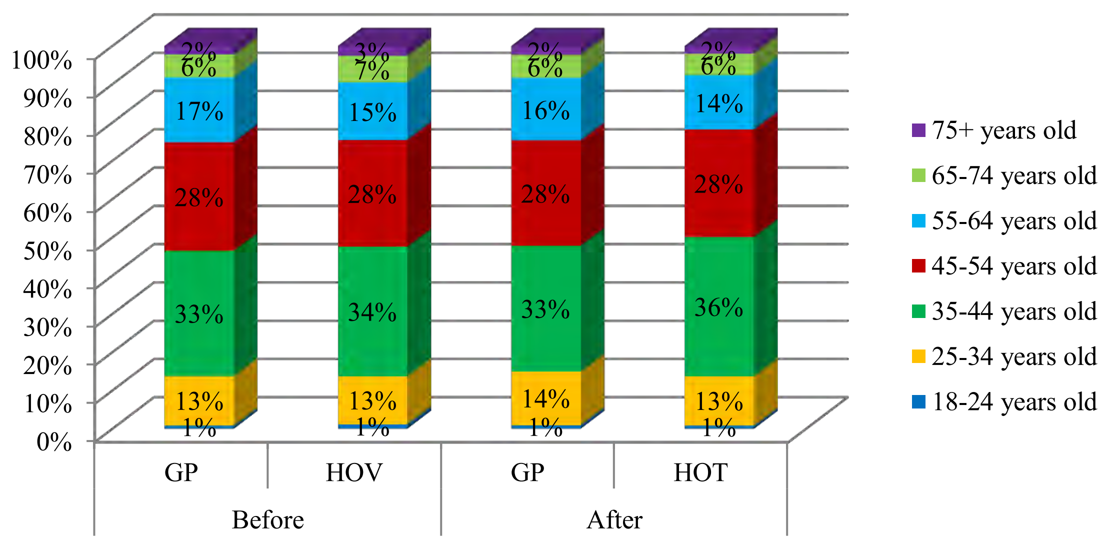


Figure 77 Head of Household Age Profiles across the Lanes (N= 241,155)

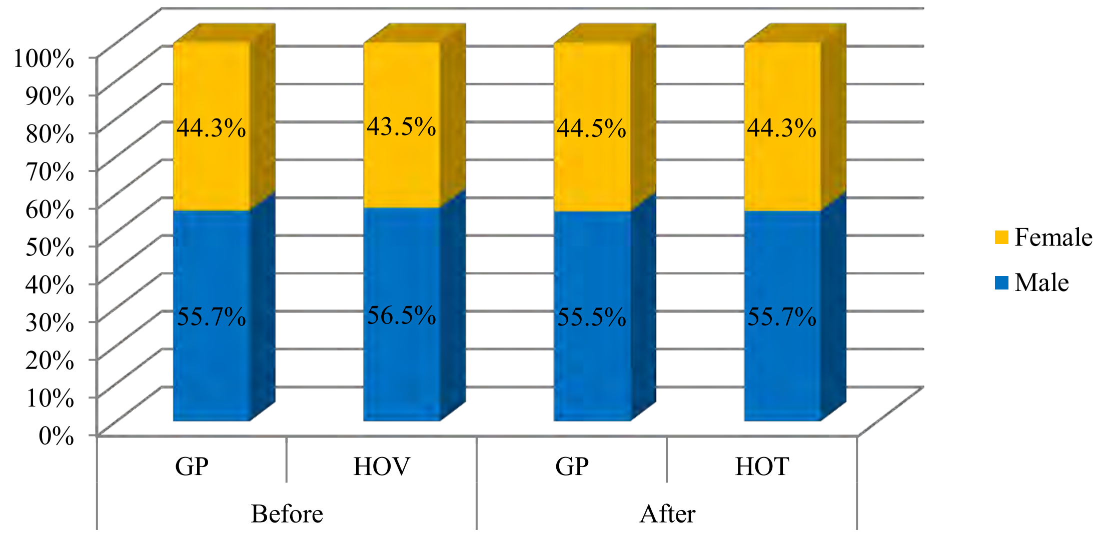


Figure 78 Head of Household Gender Profiles across the Lanes (N= 223,672)

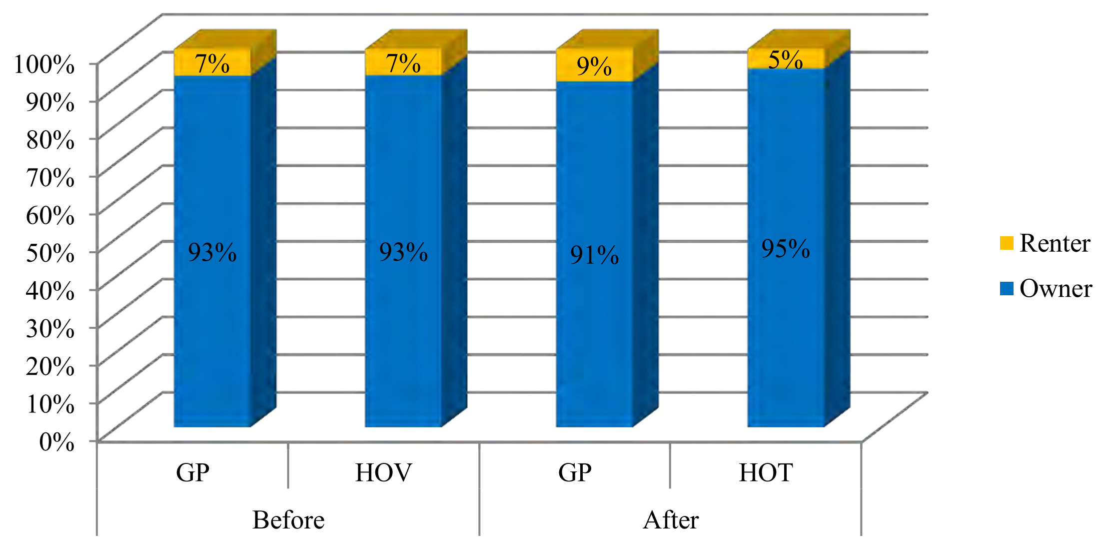


Figure 79 Household Home Ownership Profiles across the Lanes (N= 238,211)

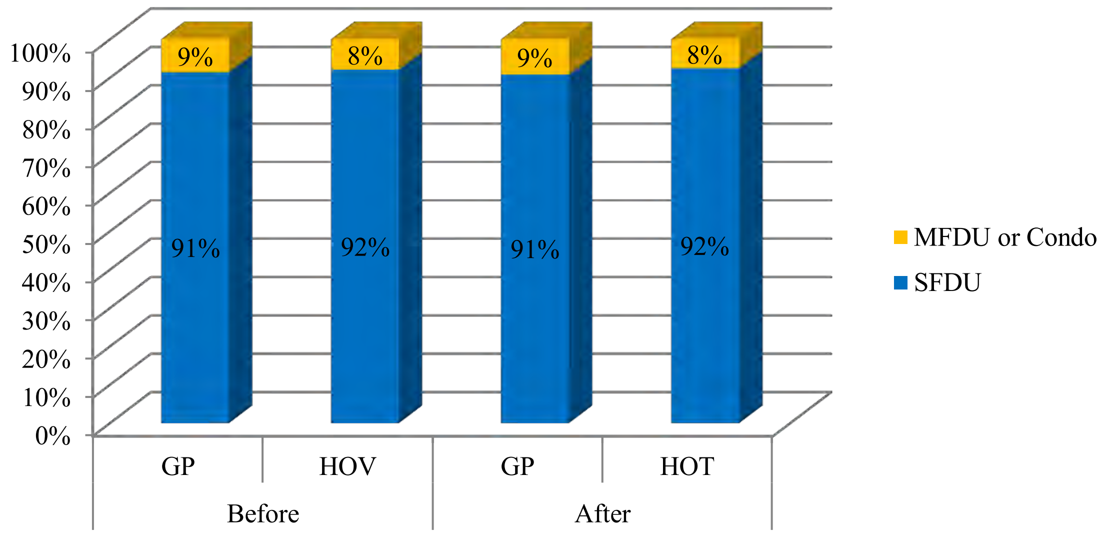


Figure 80 Households Dwelling Type Profiles across the Lanes (N= 237,776)

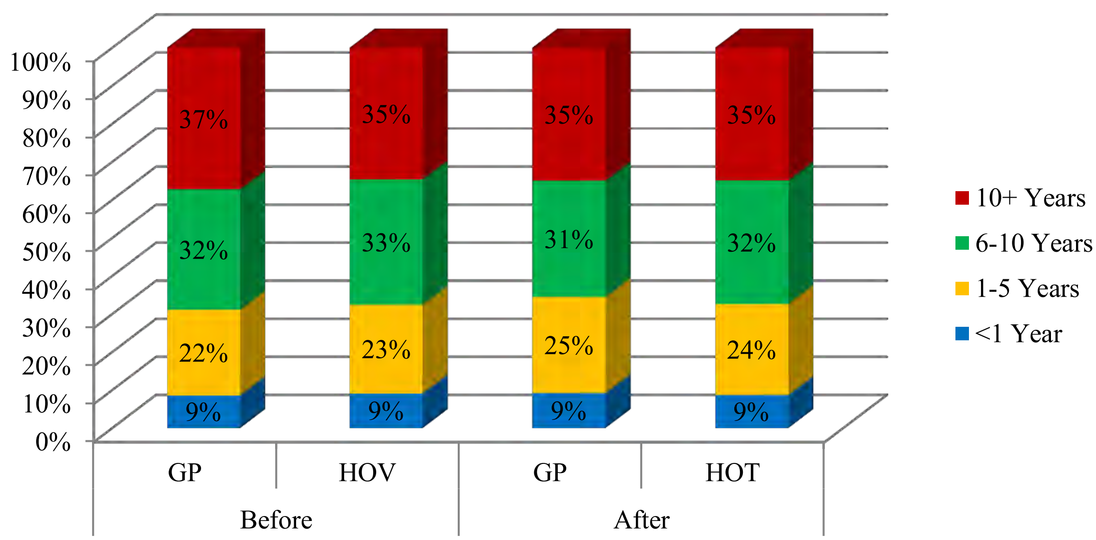


Figure 81 Households Length of Residence Profiles across the Lanes (N=241,155)

7.3. Generalized Linear Analysis: Logistic Regression

Logistic regression is a form of generalized linear model that have been discussed before in section 5.4.2. The dependent variable in logistic regression can be categorical. It is also used to predict a binary response, based on predictor variables.

Using logistic regression, two models are developed in this section. Basically, these models are mutual comparison of the four markets developed in section 6 as a function of socioeconomic attributes. The first model only focuses on the HOV frequent users and investigates if they switched to GP or continued to use HOT (HOV market vs. HOV/HOT market). The last model only focuses on the GP users and investigates if they continue to use GP lanes or switched to the HOT lane (GP market vs. HOT market).

While the goodness of fit measure have been calculated and illustrated properly, the main purpose of these models is find out the significance and direction of socioeconomic attributes that are determining the users' choice.

Figure 82 illustrates all four market groups. The blue arrows are the frequent GP users while the green arrows are the frequent HOV users. Hence, the green solid arrow is the HOV/HOT market; the green dashed arrow is the HOV market; the blue solid arrow is the GP market; and the blue dashed arrow is the HOT market. In the following section, the probability of switch or not switch for the HOV frequent users (green arrows) will be modeled in the first model. The probability of switch or not switch for the GP frequent users (blue arrows) will be modeled as well in the second model.

These models can be used in the future congestion pricing studies independently or together. For example, in a project without any HOV lane, the second model can predict the users who are more willing to switch to the new HOT lane. Furthermore, the first model can predict the carpoolers' response to a HOV-2 to HOT-3 conversion regardless of GP users' travel behavior.

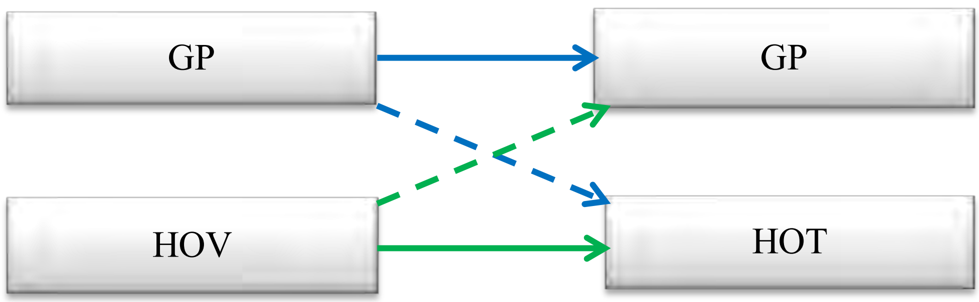


Figure 82 Four Market Groups

7.3.1. HOV users response: GP or HOT (HOV market vs. HOV/HOT market)

HOV frequent users had two options after the HOV to HOT conversion: 1. Continue to frequently use the HOT lane (HOV/HOT market); 2. Switch to the general purpose lanes (HOV market). This section is studying the impact of socioeconomic attributes on the HOV frequent users travel behavior after the conversion. The outcome of this model is very important since the only group which gets negatively impacted by the conversion is HOV-2 carpoolers who have had free access before the conversion, while have to pay toll after the conversion.

Table 26 illustrates the modeling results. The Logit 1 model includes all the variables while the Logit 2 model includes only the significant variables after step-wise elimination of insignificant variables with 95% confidence. The variables that have been eliminated –in order of elimination- are: gender, dwelling type, home ownership, length of residence, marital status, age, education, and adults. The coefficients that are not significant are colored with gray font. The full outputs for Logit models are presented in Appendix B.

In terms of goodness of fit, the model ρ^2 is 0.05 and Pseudo R^2 is 0.05. However, the Omnibus test illustrates that the model is significant compared to the constant only model with 99.999% confidence. The significant variables are income, vehicle ownership, children, and ethnicity. The Logit 2 model has been selected as the final model for further explanation. The base choice in the developed model is the "HOV/HOT market". In other words, the probability of switching to GP lanes is modeled with respect to "not switching" users as the reference category.

As it was expected, income is significantly affecting HOV users' choice of travel after the conversion. As natural log of income (divided by 1000) increases one unit, the odds of switching to general purpose lanes decreases 21%. In other words, it is more probable that higher income HOV users continue to use HOT lane while lower income HOV users switch to the general purpose lanes.

In terms of vehicle ownership, having more than one vehicle (two or three) per household decreases the odds of switching to general purpose lanes by (41% and 37% respectively) with respect to households with only one vehicle. In other words, having more than one vehicle per HOV users' households increases the chance of continuing to use the HOT lane.

Having three or more children decreases the odds of switching to general purpose lanes, compared to the households without children, by 36%. Having one and two children are not significantly affecting the users' choice. In other words, it is more probable that HOV frequent users with three or more children continue to use HOT lane compared to the other users. This was expected considering the fact that HOV-3 carpoolers continue to have free access to the HOT lane after the conversion. Other factors might also be important to make households with three or more kids pay toll and save time. Parents with three or more children should have higher willingness to save time considering their responsibilities. For example, a commuter father might want to pay toll and reach home earlier to help his wife.

Lastly, ethnicity is making the most important role in HOV frequent users travel choice after the conversion. The odds of Hispanic frequent HOV users switch to GP lanes is 237% higher than White population. In other words, the probability of switching to GP lanes by Hispanics is more than two time larger than the probability of switching by White population. Similarly, the odds of African-American frequent HOV users switch to GP lanes is 170% higher than White population. Lastly, the odds of Asian frequent HOV users switch to GP lanes is 135% higher than White population. The large significance of ethnicity in congestion pricing travel behavior studies was not expected and has not been found in any similar study.

Table 26 Logistic Regression Model Results: HOV Market vs. HOV/HOT Market (Base Choice)

Variables		Logit 1		Logit 2 (Final)	
		B	Exp(B)	B	Exp(B)
Intercept		1.372	3.942	1.466	4.330
HH Annual Income-Ln		-0.215	0.807	-0.234	0.791
Vehicles	1	0	1	0	1
	2	-.513	.598	-0.518	0.596
	3	-0.467	0.627	-0.456	0.634
	4+	-0.194	0.823	-0.171	0.842
Adults	1	0	1		
	2	0.127	1.136		
	3+	0.360	1.433		
Children	0	0	1	0	1
	1	-0.198	0.820	-0.144	0.866
	2	-0.020	0.980	0.055	1.057
	3+	-0.512	0.600	-0.450	0.638
Ethnicity	White	0	1	0	1
	African	0.504	1.656	0.535	1.707
	Hispanic	0.830	2.293	0.863	2.371
	Asian	0.291	1.338	0.300	1.350
	Other	-0.001	0.999	-0.004	0.996
Education	HS-	0	1		
	College	0.171	1.186		
	BS+	-0.023	0.977		
Marital Status	Married	0	1		
	Single	0.097	1.102		
Age	18-34 yrs	0	1		
	35-54 yrs	-0.144	0.866		
	55+ yrs	-0.239	0.787		
Gender	Male	0	1		
	Female	0.013	1.013		
Home Ownership	Renter	0	1		
	Owner	-0.134	0.874		
Dwelling Type	SFDU	0	1		
	MFDU	-0.082	0.922		
Length of Residence		0.019	1.019		
LL		-987		-991	
LL Null		-1042		-1042	
ρ^2		0.052		0.048	
Deviance		1906.673		1915	
Deviance Null		2015.819		2015	
Pseudo R^2		0.0541		0.050	
Model Significance		0.000		0.000	

7.3.2. GP users response: GP or HOT (GP market vs. HOT market)

GP frequent users (people who were not HOV frequent user) had two options after the HOV to HOT conversion: 1. Continue to use the GP lanes (GP market); 2. Switch to the HOT lane and pay for travel time saving and reliability (HOT market) or increase their vehicle occupancy to three or more. This section is studying the impact of socioeconomic attributes on the general purpose lanes' users travel behavior after the conversion.

Table 27 illustrates the modeling results. The coefficients and odds ratios of Logit model are only displayed. The Logit 1 model includes all the variables while the Logit 2 model includes only the significant variables after step-wise elimination of insignificant variables with 95% confidence. The variables that have been eliminated –in order of elimination- are: marital status, adults, dwelling type, children, and vehicles. The coefficients that are not significant are colored with gray font. The full outputs for Logit models are presented in Appendix B.

In terms of goodness of fit this model is much better than the previous models. The model ρ^2 is 0.30 and Pseudo R^2 is 0.32. Socioeconomic variables were more effective in explaining general purpose lanes' users response to pricing compared to HOV lane users. This section is trying to look at the role of each variable across the models and explains the positive and negative relationships.

The Omnibus test shows the model is significant compared to the constant only model with 99.999% confidence. The significant variables are income, ethnicity, education, age, home ownership, gender and length of residence. The Logit 2 model has been selected as the final model for further explanation. The base choice in the developed model is the “GP market”. In other word, the probability of switching to the HOT lane compared to remaining in the GP lanes has been modeled.

Similar to previous model, income is significant. The odds of GP users switch to HOT lane increases by 67% as natural logarithm of income (divided by 1000) increases by one unit. Contrary to the expectation vehicle ownership and household size variables (number of children and number of adults) are not significantly impacting GP users’ choice between HOT and general purpose lanes. The insignificance of household size variable can be explained by very low rate of non-toll trips (6%) along the HOT corridor. Therefore, the majority of the switchers to HOT lane are paying toll to use the lane; the large impact of income on probability of switching coordinates as well.

In terms of ethnicity, the odds of Hispanic population switching to HOT lane is 37% lower than White population. Similarly, the odds of Asian population switching to HOT lane is 24% lower than White population. And the odds of African-American population switching to HOT lane is 15% lower than White population.

In terms of level of education, having Bachelor or higher degree increases the odds of switching to the HOT lane 8% compared to the users with high school degree degree. In terms of age, the odds of households in older age group (head of household age: 55+ years old) switch to HOT lane is 14% lower compared to the younger age group (head of household age: 18-34 years old).

Table 27 Logistic Regression Model Results: GP Market (Base Choice) vs. HOT Market

Variables		Logit 1		Logit 2 (Final)	
		B	Exp(B)	B	Exp(B)
Intercept		-2.703	0.067	-2.844	0.058
HH Annual Income-Ln		0.478	1.613	0.502	1.652
Vehicles	1	0	1		
	2	0.158	1.171		
	3	0.143	1.153		
	4+	0.122	1.130		
Adults	1	0	1		
	2	0.024	1.024		
	3+	-0.057	0.944		
Children	0	0	1		
	1	0.086	1.090		
	2	-0.024	0.976		
	3+	0.125	1.133		
Ethnicity	White	0	1	0	1
	African	-0.149	0.862	-0.163	0.849
	Hispanic	-0.468	0.626	-0.466	0.627
	Asian	-0.263	0.769	-0.274	0.760
	Other	0.030	1.030	0.031	1.032
Education	HS-	0	1	0	1
	College	-0.067	0.935	-0.079	0.924
	BS+	0.099	1.104	0.081	1.085
Marital Status	Married	0	1		
	Single	0.001	1.001		
Age	18-34	0	1	0	1
	35-54	0.017	1.017	0.028	1.029
	55+ yrs	-0.114	0.892	-0.149	0.862
Gender	Male	0	1	0	1
	Female	0.105	1.110	0.097	1.102
Home Ownership	Renter	0	1	0	1
	Owner	0.329	1.389	0.425	1.529
Dwelling Type	SFDU	0	1		
	MFDU	-0.119	0.888		
Length of Residence		-0.101	0.904	-0.095	0.909
LL		-3713		-3718	
LL Null		-5276		-5276	
ρ^2		0.296		0.295	
Deviance		6527		6540	
Deviance Null		9653		9653	
Pseudo R^2		0.323		0.323	
Model Significance		0.000		0.000	

In terms of home ownership, the odds of owners switch to HOT lane is 53% higher compared to the renters. In terms of gender, the odds of switching by households with female head of household is 10% higher than households with male head of household. This is intuitive considering the higher level of responsibilities by female head of households. Lastly, length of residence is negatively correlated with the probability of switching to the HOT lane. As length of residence increases, the odds of switching to the HOT lane decreases by 9%.

7.3.3. Logistic Regression Modeling Discussion

Comparing across the models the two variables that are mutually significant are income and ethnicity. In terms of income, one unit increase in natural log of income (divided by 1000) decreases the odds of switching to GP lanes by 21% (vs. remaining in the HOT lane), among HOV frequent users. Among the frequent GP users, one unit increase in natural log of income (divided by 1000) increases the odds of switching to HOT lane (HOT market) by 67%. Considering the HOT lane toll payment requirement for single and double occupant vehicles, the impact of income is intuitively explanatory.

The relationship between income (divided by 1000) and its natural log is illustrated in Figure 83. The application of natural log transformation of income increase the sensitivity of the model to lower income population compared to higher income. For instance, by one unit increase in natural log of income (divided by 1000) the odds of switching to HOT lane increases 67% for GP frequent users. This one unit corresponds to increase in household income from $20,000 to $40,000; similarly, it corresponds to increase in household income from $60,000 to $150,000 as well. In fact, the same amount of difference in income is more important for lower income households compared to the higher income households. This fact has been considered in the model using natural log transformation of income.

The other critical issue with respect to the income is the fact that the source of income variable by marketing data is unknown (self-reported data is not the basis for all the income values). Even, considering only self-reported sources, people may not report their total earnings per year. Especially, higher income people may only report their salaries, while they have other income sources such as interests. Moreover, regardless of income people wealth are also different. A low income household may have high willingness to pay considering large amount of wealth. Although there are all sorts of inaccuracies with respect to the income variable, the fact that sensitivity of the model to income decreases as income increases is becoming more reasonable.

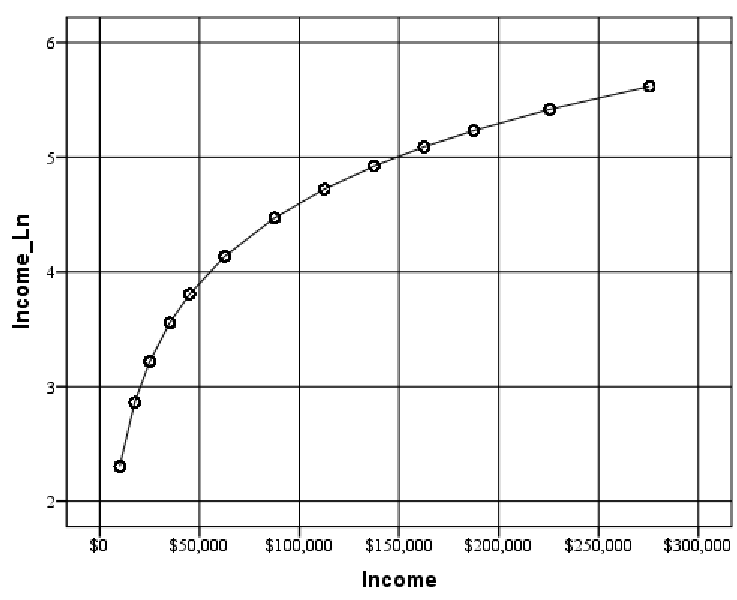


Figure 83 Relationship between Original Income Value (divided by 1000) and Natural Log Transformation

Ethnic groups have become significantly determinant variables in users travel choice before and after the conversion sometimes even more than income. For example, the probability of switching to GP lanes by Hispanic frequent HOV users is 2.3 times larger than White frequent HOV users. Moreover, the probability of switching to HOT lane by Hispanic frequent GP users is 37% less than White frequent GP users. Asian and African-American populations show similar results with lower impact.

It has been previously explained that the correlation between socioeconomic attributes (presented or not presented in the model) could potentially help explain the presence of ethnicity in the developed models. Furthermore, the fact that use of the HOT lane (whether paying a toll or not) requires the possession of a Peachpass (RFID transponder installed on the windshield) and a linked bank account, the act of registering for lane use may also be a potential underlying reason for lower HOT lane preference across ethnic groups. Before drawing a conclusion that ethnicity is the casual variable in play, additional research is warranted, including stated preference surveys.

7.4. Generalized Linear Analysis: Count Model

This section of the study applies generalized linear count models to investigate the role of socioeconomic variables on number of observation per household along the HOV and HOT lanes. Count models, developed in this section, can be used to forecast managed lane use frequency for future Traffic and Revenue studies. The count models are a form of generalized linear models for which the dependent variable is in the count format: 0, 1, 2, 3,… . Count models, in this section, are trying to study the impact of socioeconomic variables on number of times a household have been observed either in HOV or HOT lane controlling for total number of times the household used the corridor.

Hence, the dependent variables are frequency of observation along HOV lane and HOT lane.

Poisson and negative binomial regression models are designed to analyze count data. However, Poisson and negative binomial models differ in regard to their assumptions of the conditional mean and variance of the dependent variable. Poisson models assume that the conditional mean and variance of the distribution are equal. Negative binomial regression models do not assume an equal mean and variance and particularly correct for over dispersion in the data, which is when the variance is greater than the conditional mean.

Looking at managed lanes use observation frequency data (Table 19), the variance of the HOV lane use frequency (2.93) is significantly larger than the mean (0.35). Similarly, the variance of the HOT lane use frequency (2.00) is significantly larger than the mean (0.24). Therefore, it is more logical to use the negative binomial distribution. However, Poisson distribution are far from nonexistent, with some researchers even observing the presence of both Poisson and negative binomial within the same study (Braga and Bond, 2008). In this study we have estimated both of the models which have resulted in very similar goodness of fit and coefficients. However, the models with the negative binomial distribution have been illustrated in the dissertation and used for the explanation. The previously explained goodness of fit measures, same as logistic regression, have been used here. However, instead of odds ratios count models coefficients are explained by Incident Rate Ratios (IRR). IRR explains the rate of change in number of counts by a unit increase in an exposure variable (for scale variables) or by presence of an exposure variable relative to a reference category (for categorical variables). Because Negative Binomial model also uses Log link, IRR is equal to anti-log of coefficients (Exp (B)), similar to odds ratios.

7.4.1. HOV Lane Use Count Model

The Negative Binomial count model has been developed to model households' HOV lane frequency of use including all the socioeconomic attributes. All the variables except for marital status become significant at 99% confidence. The model Pseudo $\mathbf{R}^2$ is 0.52 and $\boldsymbol{\rho}^2$ is 0.41. Table 28 illustrates the final model fit after eliminating the insignificant variable (marital status). Income is slightly and negatively correlated with HOV lane use observation frequency. If the natural log of income (divided by 1000) increases one unit, the HOV lane use frequency decreases 3%. This implies that increase in income decrease the probability of using HOV lane slightly.

In terms of vehicle ownership, having more vehicles increases the chance of HOV trip. Having two vehicles rather than on vehicle increases the number of HOV trips 8%. Similarly, having three vehicles rather than one vehicle increases the number of HOV trips 19%. And, having four or more vehicles rather than one vehicle increases the number of HOV trips by 27%. Coordinating with the previous observation, number of HOV use frequency increases as vehicle ownership increases controlling for total corridor use and other socioeconomic variables.

Household size related variables relations to HOV use frequency are significant. Having more than one adult per household increases the number of HOV trips 10%, compared to having only one adult per household. Regarding number of children, having one or two children decreases number of HOV trips 3% and 7% respectively, with respect to households without children. Having three or more children is not significantly different from having no children. Therefore, during the peak hours having two or more adults per households is increasing the chance of carpooling while having children is not significantly affecting the carpool formation (fampooling). This result implies that households with two or more adults have the highest probability of making HOV trips.

In terms of ethnicity, African American households make 23% fewer HOV trips; and, Hispanic and Asian households make 44% and 49% more HOV trips, with respect to the White population, controlling for total corridor use and other socioeconomic attributes. Once again, the substantial impact of ethnicity on managed lane use is very interesting and has not been found in previous studies.

Education is negatively correlated with HOV lane frequency of use. Households with college degree make HOV trips 4% fewer; and, households with bachelor or higher degree make 13% fewer HOV trips compared to the households with high school degree (or lower).

In terms of head of household age, older households with head of household age of 55+ years old makes 5% more and mid-age households with head of household age of 34-55 years old make 7% more HOV trips, compared to the younger households with head of household age of 18-34 years old.

With respect to gender, households with female head of household make 5% fewer HOV trips compared to households with male head of household. In terms of home ownership, owners make HOV trips 19% more than renters. In terms of dwelling type, multi-family units/apartments/condos make 10% fewer HOV trips compared to single family units. Lastly, increase in length of residence very slightly decreases the number of HOV trips by 2%.

Table 28 HOV Lane Negative Binomial Count Model

Variables		B	Std. Error	Sig.	IRR Exp (B)	95% Wald CI for Exp (B)	
						Lower	Upper
Intercept		-2.08	0.05	0.00	0.13	0.11	0.14
Total Frequency		0.05	0.00	0.00	1.05	1.05	1.05
HH Annual Income-Ln		-0.03	0.01	0.01	0.97	0.94	0.99
Vehicles	1	0	.	.	1.00	.	.
	2	0.08	0.02	0.00	1.08	1.04	1.12
	3	0.17	0.02	0.00	1.19	1.14	1.24
	4+	0.24	0.02	0.00	1.27	1.22	1.32
Adults	1	0	.	.	1.00	.	.
	2	0.10	0.02	0.00	1.10	1.07	1.14
	3+	0.09	0.02	0.00	1.09	1.05	1.14
Children	0	0	.	.	1.00	.	.
	1	-0.03	0.02	0.07	0.97	0.93	1.00
	2	-0.07	0.03	0.00	0.93	0.88	0.98
	3+	-0.01	0.02	0.68	0.99	0.94	1.04
Ethnicity	White	0	.	.	1.00	.	.
	African	-0.26	0.02	0.00	0.77	0.74	0.81
	Hispanic	0.37	0.02	0.00	1.44	1.38	1.50
	Asian	0.40	0.02	0.00	1.49	1.42	1.56
	Other	0.07	0.03	0.01	1.07	1.02	1.13
Education	HS-	0	.	.	1.00	.	.
	College	-0.04	0.02	0.03	0.96	0.93	1.00
	BS+	-0.14	0.02	0.00	0.87	0.84	0.90
Age	18-34 yrs	0	.	.	1.00	.	.
	35-54 yrs	0.06	0.02	0.00	1.07	1.03	1.11
	55+ yrs	0.05	0.02	0.02	1.05	1.01	1.10
Gender	Male	0	.	.	1.00	.	.
	Female	-0.05	0.01	0.00	0.95	0.93	0.98
Home Ownership	Renter	0	.	.	1.00	.	.
	Owner	0.17	0.03	0.00	1.19	1.13	1.25
Dwelling Type	SFDU	0	.	.	1.00	.	.
	MFDU	-0.10	0.03	0.00	0.90	0.85	0.95
Length of Residence		-0.02	0.00	0.00	0.98	0.97	0.99
Log Likelihood		-85,663					
Deviance		107,218					
AIC		171,372					
ρ^2		0.41					
Pseudo R^2		0.53					
Model Significance		0.000					

7.4.2. HOT Lane Use Count Model

Similar to HOV lane use frequency, the Negative Binomial model has been developed for HOT lane use frequency. All the variables except for gender become significant at 95% confidence. The model Pseudo $\mathbf{R}^2$ is 0.67 and $\boldsymbol{\rho}^2$ is 0.60. Considering the goodness of fit, socioeconomic variables are more powerful in modeling HOT lane use frequency compared to HOV lane use frequency.

Table 29 illustrates the final model fit after eliminating the insignificant variable (gender). Income is positively correlated with HOT lane use frequency. By increasing the natural log of income (divided by 1000) by one unit, the number of HOT lane trips per household increases 37%. This is perfectly intuitive, considering the toll payment requirement for single and double occupant vehicle trips along the HOT lane, which corresponds to about 94% of the HOT trips.

In terms of vehicle ownership, having more than one vehicle per household increases the number of HOT trips compared to household with only one vehicle registered. Specifically, having two vehicles increases the number of HOT trips by 16%; having three vehicles increases the number of HOT trips by 15%; and, having four or more vehicles increases the number of HOT trips by 12% compared to households with only one vehicle registered.

In terms of household size related variables, households with two or more adult makes 12% fewer HOT trips compared to households with only one adult. Having children increases the number of HOT trips. Having one, two, and three or more children increases the number of HOT use frequency by 7%, 13%, and 22% respectively. The presence of children, not only increases the chance of carpooling, but also increases the parents' responsibilities and accordingly value of time.

In terms of ethnicity, African-American and Hispanic populations make about 28% fewer HOT trips and Asian population makes 15% fewer HOT trips, compared to White population. The findings regarding ethnicity coordinate well across all the models.

In terms of education, households with Bachelor (or higher) degree are making 5% more HOT trips compares to households with high school degree or lower, controlling for total corridor use and other socioeconomic factors.

In terms of marital status, single households make 8% fewer HOT trips compared to married households. In terms of age, older households with head of household age of 55+ years are making HOT trips 11% fewer than younger households with head of household age of less than 55 years old.

In terms of home ownership, owners make 40% more HOT trips compared to renters. Regarding dwelling type, multi-family units/apartments/condos are making 9% fewer HOT trips compared to single family units. Lastly, number of HOT lane trips decreases 7% by increasing length of residence.

Table 29 HOT Lane Negative Binomial Count Model

Variables		B	Std. Error	Sig.	IRR Exp (B)	95% Wald CI for Exp (B)	
						Lower	Upper
Intercept		-4.02	0.08	0.00	0.02	0.02	0.02
Total Frequency		0.07	0.00	0.00	1.07	1.07	1.07
HH Annual Income-Ln		0.32	0.02	0.00	1.37	1.33	1.42
Vehicles	1	0		.	1	.	.
	2	0.15	0.02	0.00	1.16	1.11	1.22
	3	0.14	0.03	0.00	1.15	1.09	1.21
	4+	0.11	0.03	0.00	1.12	1.06	1.18
Adults	1	0		.	1	.	.
	2	-0.14	0.03	0.00	0.87	0.82	0.92
	3+	-0.13	0.03	0.00	0.88	0.82	0.94
Children	0	0		.	1	.	.
	1	0.07	0.03	0.01	1.07	1.02	1.12
	2	0.12	0.03	0.00	1.13	1.06	1.21
	3+	0.20	0.03	0.00	1.22	1.15	1.30
Ethnicity	White	0		.	1	.	.
	African	-0.34	0.03	0.00	0.71	0.67	0.76
	Hispanic	-0.33	0.03	0.00	0.72	0.67	0.77
	Asian	-0.16	0.03	0.00	0.85	0.80	0.91
	Other	-0.07	0.03	0.03	0.93	0.87	0.99
Education	HS-	0		.	1	.	.
	College	-0.01	0.02	0.62	0.99	0.95	1.03
	BS+	0.05	0.02	0.04	1.05	1.00	1.10
Marital Status	Married	0		.	1	.	.
	Single	-0.08	0.03	0.01	0.92	0.87	0.98
Age	18-34 yrs	0		.	1	.	.
	35-54 yrs	0.00	0.03	0.99	1.00	0.95	1.05
	55+ yrs	-0.12	0.03	0.00	0.89	0.84	0.95
Home Ownership	Renter	0		.	1	.	.
	Owner	0.33	0.04	0.00	1.40	1.30	1.50
Dwelling Type	SFDU	0		.	1	.	.
	MFDU	0.09	0.04	0.01	1.09	1.02	1.17
Length of Residence		-0.07	0.01	0.00	0.93	0.92	0.94
Log Likelihood		-56,135					
Deviance		81,243					
AIC		112,316					
ρ^2		0.601					
Pseudo R^2		0.676					
Model Significance		0.000					

7.4.3. Count Modeling Discussion

The fact that the goodness of fit of the HOT lane model is substantially better than HOV lane model implies more association of travel behavior to socioeconomic attributes in presence of a pricing strategy.

With respect to the HOV model, ethnicity (Hispanic and Asian) has the highest positive impact (deviation of the IRRs from one is the highest), followed by vehicle ownership and home ownership; however, ethnicity (African-American) followed by higher education (BS or more) have the highest negative impact.

With respect to the HOT model, income and home ownership have the highest positive impact followed by number of children and vehicle ownership; however, ethnicity (African-American, Hispanic, and Asian) and number of adults have the highest negative impacts. Therefore, the impact of ethnicity is very substantial followed by income, home ownership, vehicle ownership, and household-size related variables.

While ethnicity is a significantly correlated with HOV and HOT lane use, other latent life style variables that are also correlated with ethnicity might explain the observation. For example, carpooling may be more prevalent for clustered households with similar socio-demographic characteristics that work at proximal jobs sites. Hence the higher presence of Hispanic households in the carpooling category may not be related to ethnicity at all, but rather the more specific characteristics of specific spatial ethnic clusters.

Moreover, there are certainly correlations between ethnicity and other socioeconomic attributes. For example, high income (more than $120,000) White households were 2.8 times more prevalent than their Hispanic counterparts. Therefore, income may show up as a general non-linear variable as well as an interaction variable with ethnicity or other parameters. Application of interactive variables could potentially help us better understand the independent impact of socioeconomic attributes. For example, it will be interesting to develop the models across different income categories. Using this, we can expand our knowledge about users travel behavior to assess whether low income users from different ethnic group respond differently to congestion pricing. Considering the decent amount of work already conducted in this study, further expansion of the developed models, using interactive variables will follow the completion of this work.

CHAPTER 8

SPATIAL ANALYSIS

The concept of a “commutershed”, which is also sometimes referred to as a “catchment area”, is regularly employed by researchers to estimate facility travel demand (Horner and Groves, 2007). Commutersheds can be developed for any transportation facility such as highways, transit routes, and park and ride facilities to identify the potential users of the corridor. As in other travel behavior studies, surveying has been the main method of commutershed data collection. However, the small sample size of typical survey methods limits the possibility of conducting GIS analysis.

This study utilizes over one million collected license plates over a two-year study period to define a very detailed commutershed for the I-85 HOV-to-HOT corridor. The comparison of the before and after the conversion commutersheds helps researchers better understand whether the HOT lane significantly impacted the spatial distribution of corridor users. Because the HOT lane provides lower and reliable travel times, it is expected that people living farther away from the commutershed before conversion may start to use the corridor. Although expanding the service area of a transportation facility is a potential mobility advantage, the potential disadvantages such as “urban sprawl” should also be acknowledged.

8.1. Commutershed Visualization and Analysis

8.1.1. Density Maps

Among all the vehicles observed over seven seasons of data collection (180 two-hour sessions) along I-85 HOV-to-HOT corridor, 87% of the vehicle’s registration addresses belong to Atlanta metro area are predominantly distributed around the HOV-HOT corridor. Because of the very large size of collected data, GIS (Geographic Information Systems) analysis tools have been used to display and analyze the spatial distribution of the license plate registration location.

First, the ArcGIS Point Density Function has been used to develop corridor commutershed maps. The Point Density Function calculates the density of point features around each output raster cell. Conceptually, a neighborhood is defined around each raster cell center, and the number of points that fall within the neighborhood is totaled and divided by the area of the neighborhood. The population field is used to weights the observation data. The weights have been used to differentiate between households with different numbers of observations. Accordingly, the weights for the households observed more frequently in the corridor are greater, relative to the households observed less frequently. The study area consists of 311 columns and 251 rows. The cell size is 1782 feet. The size of the cell is automatically selected by the software. The total study area is 8,891 square miles. However, not all of the cells have a density value.

Figure 84 shows the point density commutershed map for all lanes on the corridor (HOV lane and all general purpose lanes) before the conversion. Figure 85 shows the same map for after the conversion (HOT lane and all general purpose lanes).

Considering the larger number of data collection sessions before the conversion (about 30% larger), the larger density values are expected on the before conversion commutershed maps. The largest calculated observation density per square mile is 3,935 for before the conversion compared to 2,221 for after the conversion. To compare the figures visually, the density values have been classified to 20 quantile classes and the same color ramp has been used. Therefore similar colors have relatively same density comparing between the commutershed maps, although they do not have exact same density values.

Comparing all the lanes commutershed maps, before and after the conversion, the overall spatial pattern of the commutershed has not changed visually to a great extent. The exact amount and location of the change will be further investigated in the forthcoming sections.

Figure 86 illustrates the point density commutershed for all the general purpose lanes before the conversion. Figure 87 illustrates the same map for after the conversion. Similarly, comparing general purpose lanes commutershed maps before and after the conversion, the overall spatial pattern of the commutershed has not changed visually to a great extent.

Figure 88 illustrates the HOV lane point density commutershed map. Figure 89 illustrates the HOT lane point density commutershed map. Considering the pricing scheme (for travel time saving and reliability) of the conversion along the HOT corridor, a significant change (mainly expansion) in the commutershed was expected. However, the HOT lane commutershed has generally retracted, while locally expanded in some areas. The exact amount and location of the change will be further investigated in the forthcoming sections.

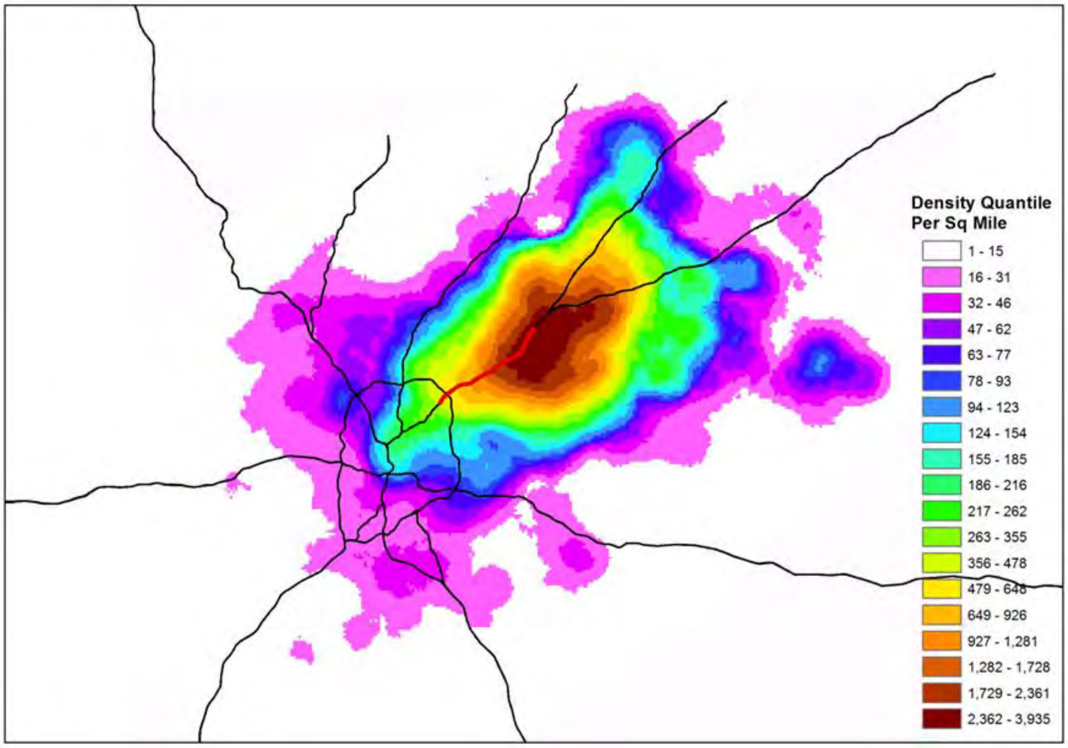


Figure 84 I-85 Corridor Before Conversion Commutershed

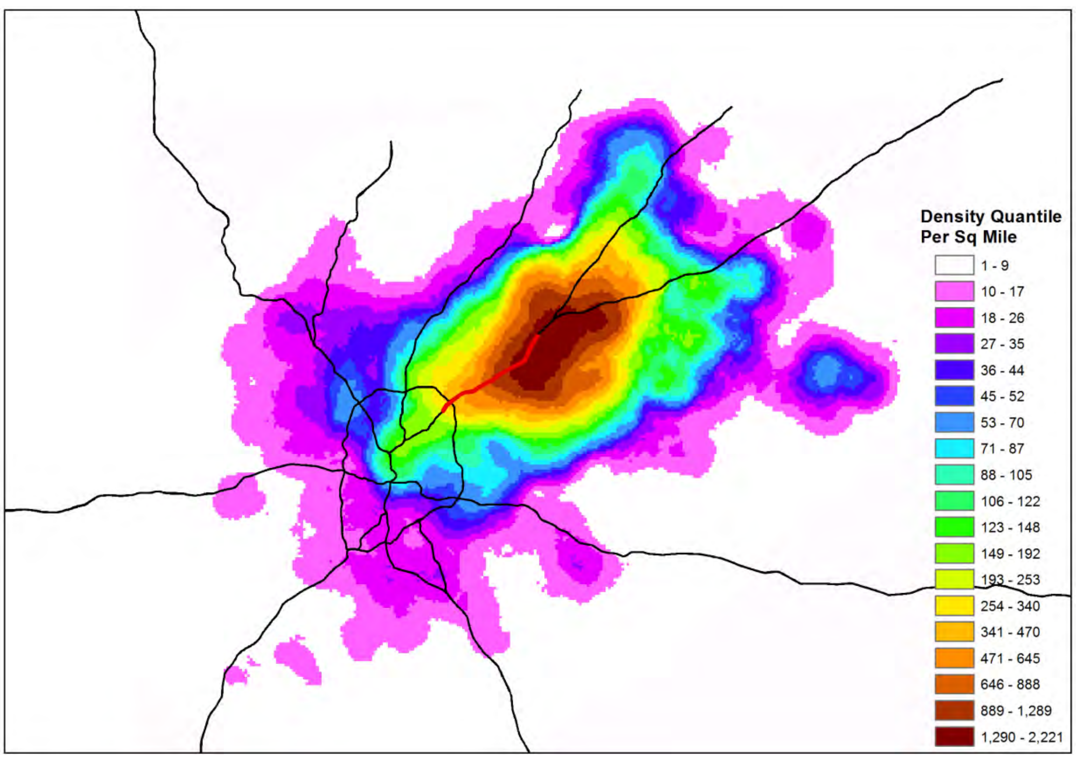


Figure 85 I-85 Corridor After Conversion Commutershed

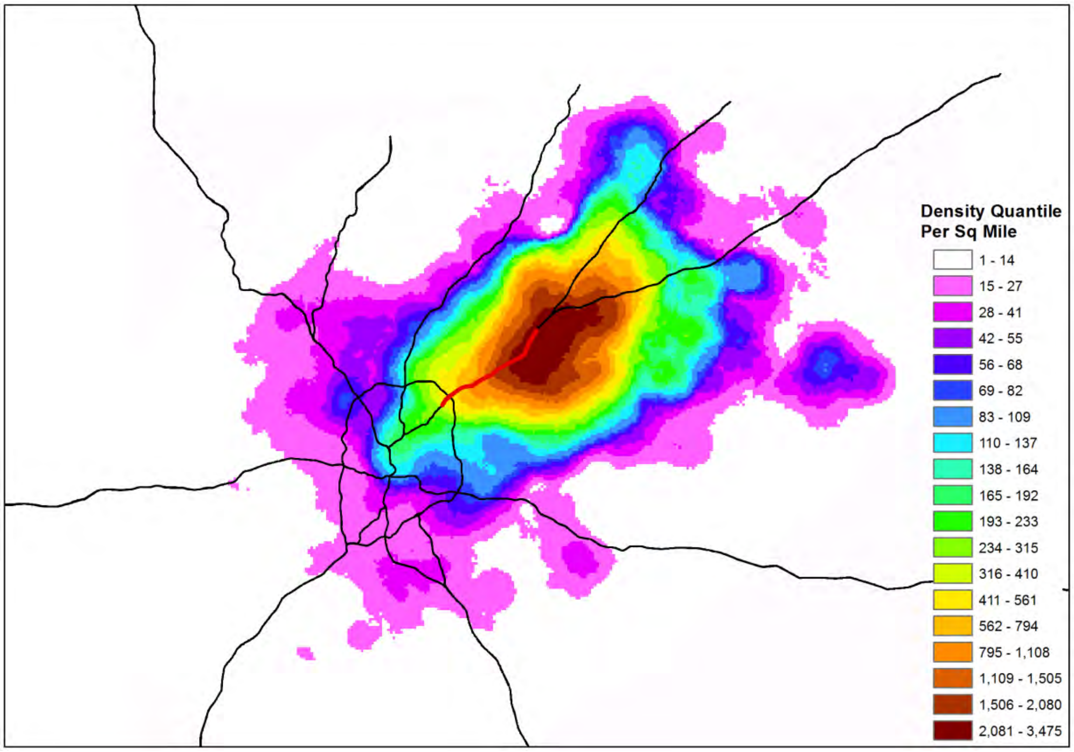


Figure 86 I-85 Corridor General Purpose Lanes Before Conversion Commutershed

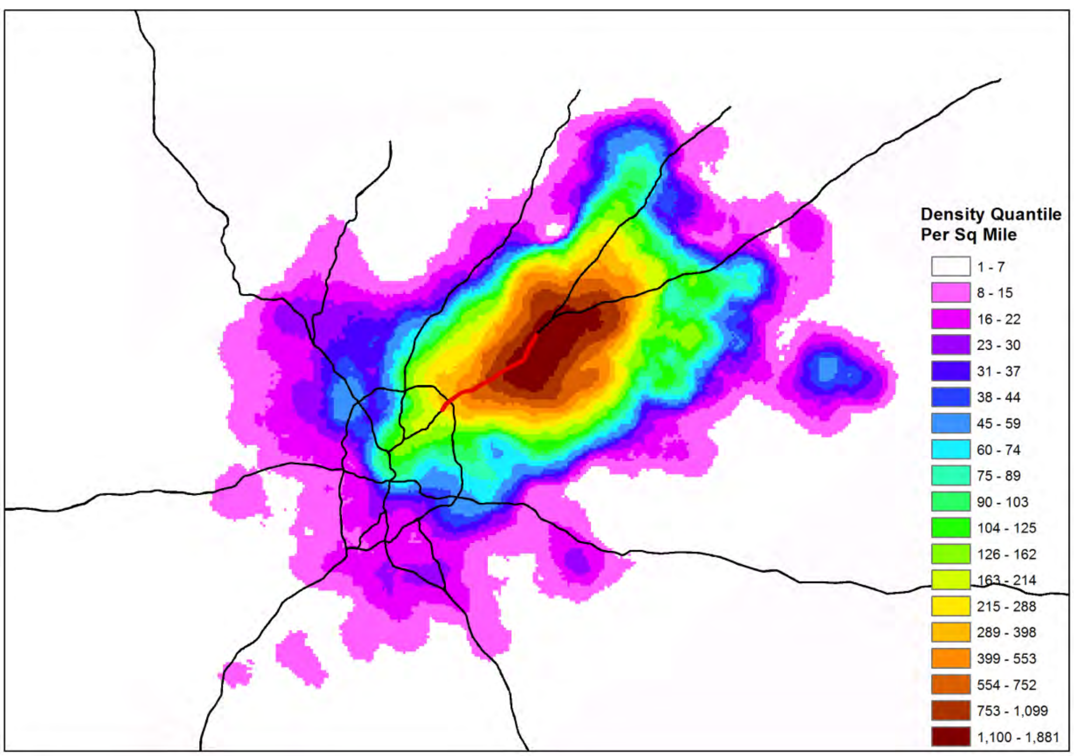


Figure 87 I-85 Corridor General Purpose Lanes After Conversion Commutershed

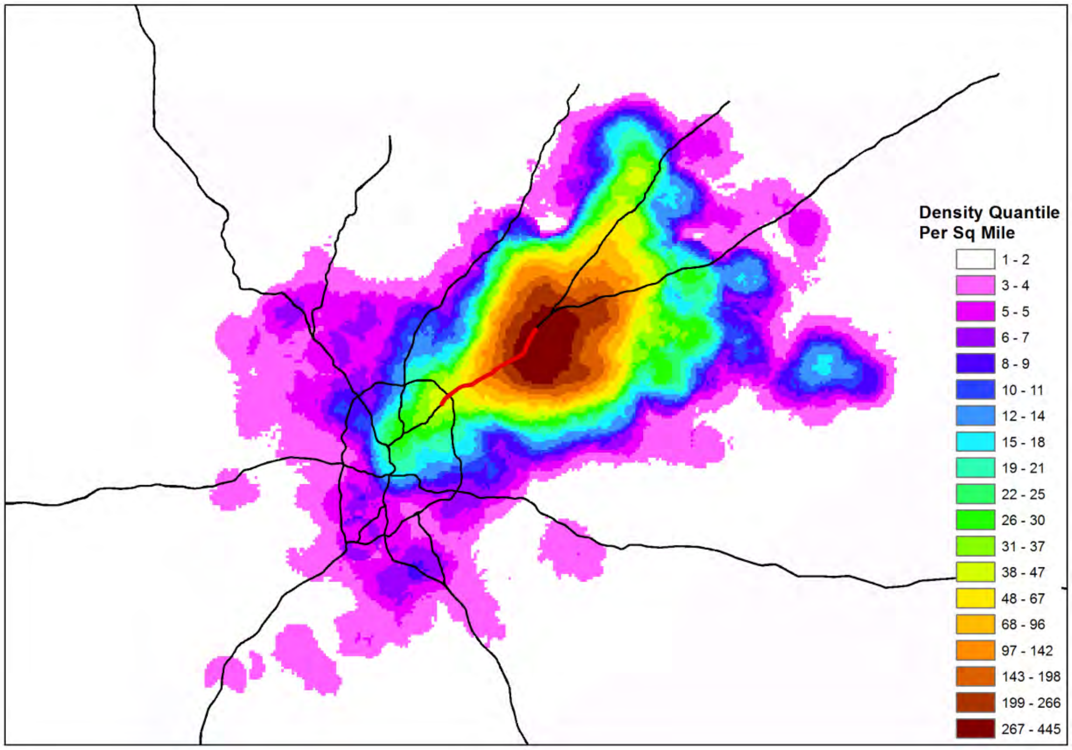


Figure 88 I-85 Corridor HOV Lane Commutershed

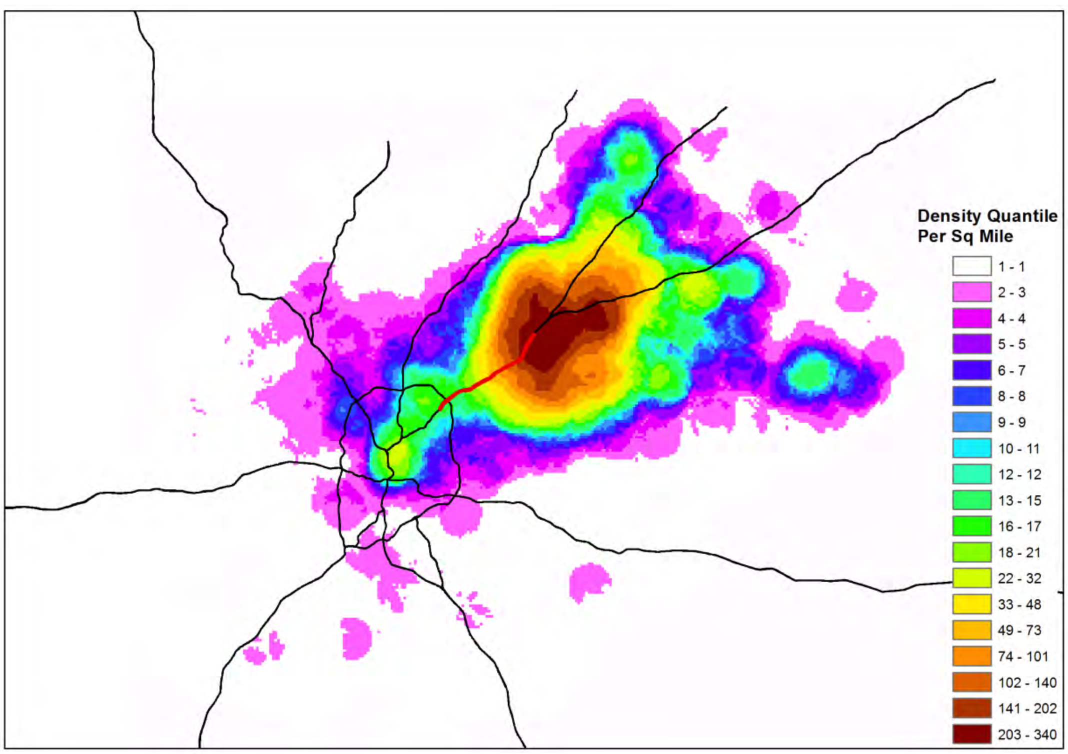


Figure 89 I-85 Corridor HOT Lane Commutershed

8.1.2. Directional Distribution

The second method of spatial distribution analysis is Directional Distribution Ellipse. The ellipse is referred to as the standard deviational ellipse, because the method calculates the standard deviation of the x coordinates and y coordinates from the mean center to define the axes of the ellipse. For example, ellipses developed based on two standard distribution cover 95% of the observations in the map. Also, the ellipse allows you to see if the distribution of features is elongated and hence has a particular orientation. Similarly, the ellipses for the general purpose lanes (before and after the conversion), HOV lane and HOT lane have been developed and presented in Figure 90. The ellipses have been generated based on two standard deviations and therefore include 95% of the observed commuters.

Table 30 illustrates the ellipses metrics. The HOT lane commutershed retracted and the center point of the HOT commutershed is slightly displaced to northeast corner compared to the HOV commutershed. The length of displacement is about one mile to the north and one mile to the east. The HOT lane ellipse is also rotated about six degrees (clockwise) to the east. Therefor the HOT lane commutershed is retracted more in the north-south direction (about three miles) compared to east-west direction (about one and half mile). Consequently, the HOT lane ellipse area is 16% smaller than the HOV lane. These changes indicate that there has been a change in the managed lane user market.

Considering the fact that the data for this study were only collected during peak hours, the observed decrease and displacement in the HOT lane commutershed area after the conversion might only be associated with commuting trips during peak-hour.

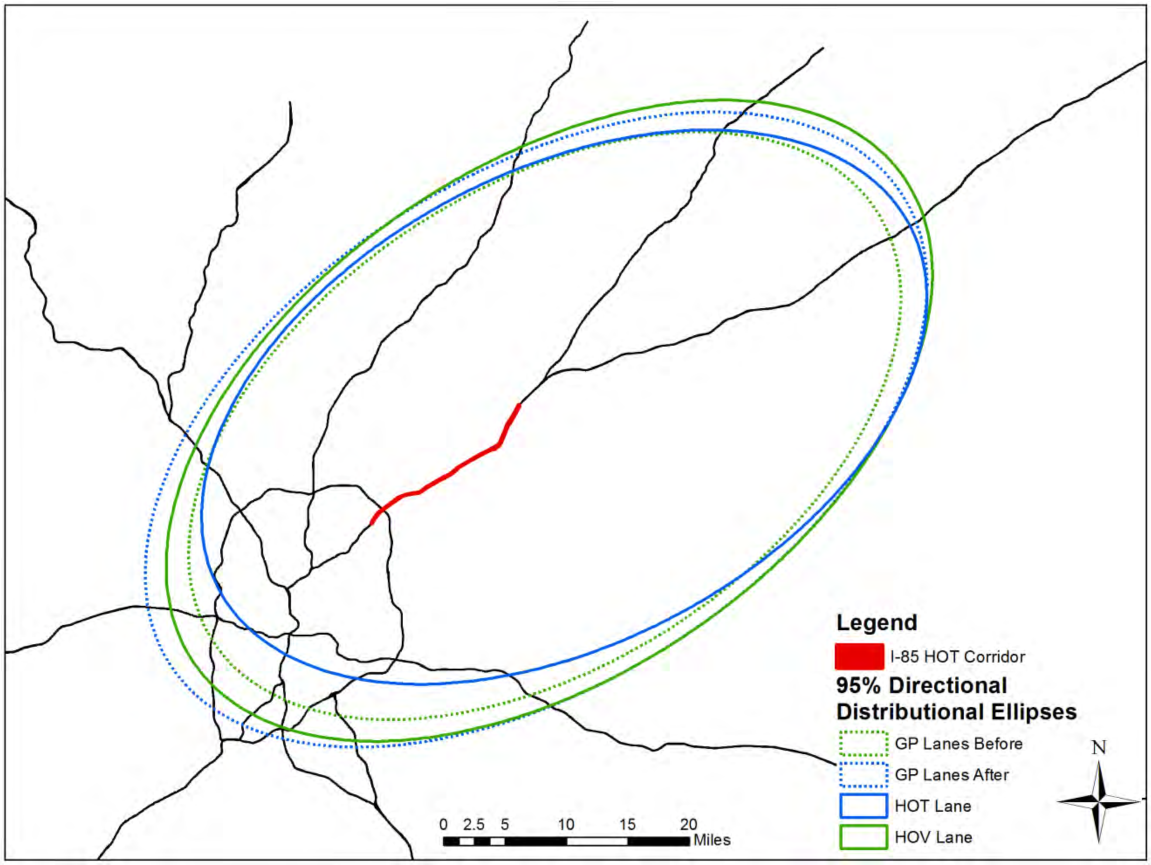


Figure 90 Directional Distribution Ellipses

In contrast, the general purpose lane commutershed expanded in both directions after the conversion. The center of the general purpose lanes ellipse has displaced very slightly (less than a mile) toward the southwest (opposite to the direction of displacement in the managed lane commutershed). The ellipse rotation angle has changed less than a degree. The general purpose lanes commutershed has been expanded more than one mile along the east-west direction and more than three mile along the north-south direction. Consequently, the area of the general purpose lanes ellipse increased 17%.

Table 30 Directional Distributional Ellipses Metrics (Unit: mile)

	Center		**SD Dis.**		**Rotation Angle**	**Area**
	(X)	(Y)	(X)	(Y)		
All Lanes (Before)	442.86	274.98	19.51	32.50	56.82	1992.05
All Lanes (After)	442.40	274.86	20.48	35.04	57.77	2254.32
GP Lanes (Before)	442.81	274.93	19.36	32.16	57.00	1956.02
GP Lanes (After)	442.11	274.63	20.64	35.38	57.33	2293.78
HOV Lane	443.18	275.35	20.67	34.98	55.71	2271.27
HOT Lane	444.40	276.46	19.22	31.76	62.26	1917.29
HOT vs. HOV	1.22	1.11	-1.45	-3.22	6.55	-16%
GP Lanes After vs. Before	-0.70	-0.30	1.28	3.22	0.33	17%
All Lanes After vs. Before	-0.45	-0.12	0.97	2.54	0.95	13%

Considering all of the I-85 corridor lanes, a slight displacement has occurred toward south-west (less than one mile) after the conversion. The rotation angle has moved by about one degree, clockwise to the east. The corridor commutershed has expanded two and half miles along the north-south direction and a little less than one mile along the east-west direction. Consequently, the corridor commutershed area has expanded in catchment area by about 13%.

Because of the observed retraction in managed lane commutershed and observed expansion in general purpose lanes commutershed, it seems reasonable to conclude that some of the more remotely located HOV lane users have switched to general purpose lanes or changed their commute time or route. Moreover, the overall expansion in the corridor commutershed (considering the HOT lane commutershed retraction) implies an overall increase in service area of the corridor, after providing priced HOT lane (which provides travel time saving and reliability option for the managed lane).

Lastly, directional distributional analysis is the best method to evaluate the overall directional displacement of the commutershed. However, it has the disadvantage of including the areas that are not actually part of the real commutershed in the analysis. Therefore, the estimation error in area calculation is unavoidable. The method applied in the next section will help resolve this problem and produces a more accurate estimate of the change in the area of the commutershed.

8.1.3. Fuzzy Membership

Considering the different amount of data collection before and after the conversion, the mutual quantitative comparison of the real commutershed density values is not feasible. Therefore, a linear transformation is performed on the raw density values to be able to compare the output values mutually. The "Fuzzy Membership" function in ArcGIS transforms the input raster values to a 0 to 1 scale, indicating the strength of a membership in a set, based on a fuzzification algorithm. In this case a linear algorithm, which gives the highest value 1 and the lowest value 0, has been used. A value of one indicates full membership (in this case the absolute corridor commuter), with membership decreasing to zero, indicating it is not a member of the fuzzy set (ESRI, 2013).

Figure 91 illustrates the fuzzy membership outputs of the six developed commutershed maps. The dark blue cells illustrate close to full membership while the light blue cells illustrate no membership. Each cell has a fuzzy value between zero and one. The left column figures show the before conversion commutershed and the right column figures show the after conversion commutersheds. The I-85 HOV-to-HOT corridor is highlighted in red.

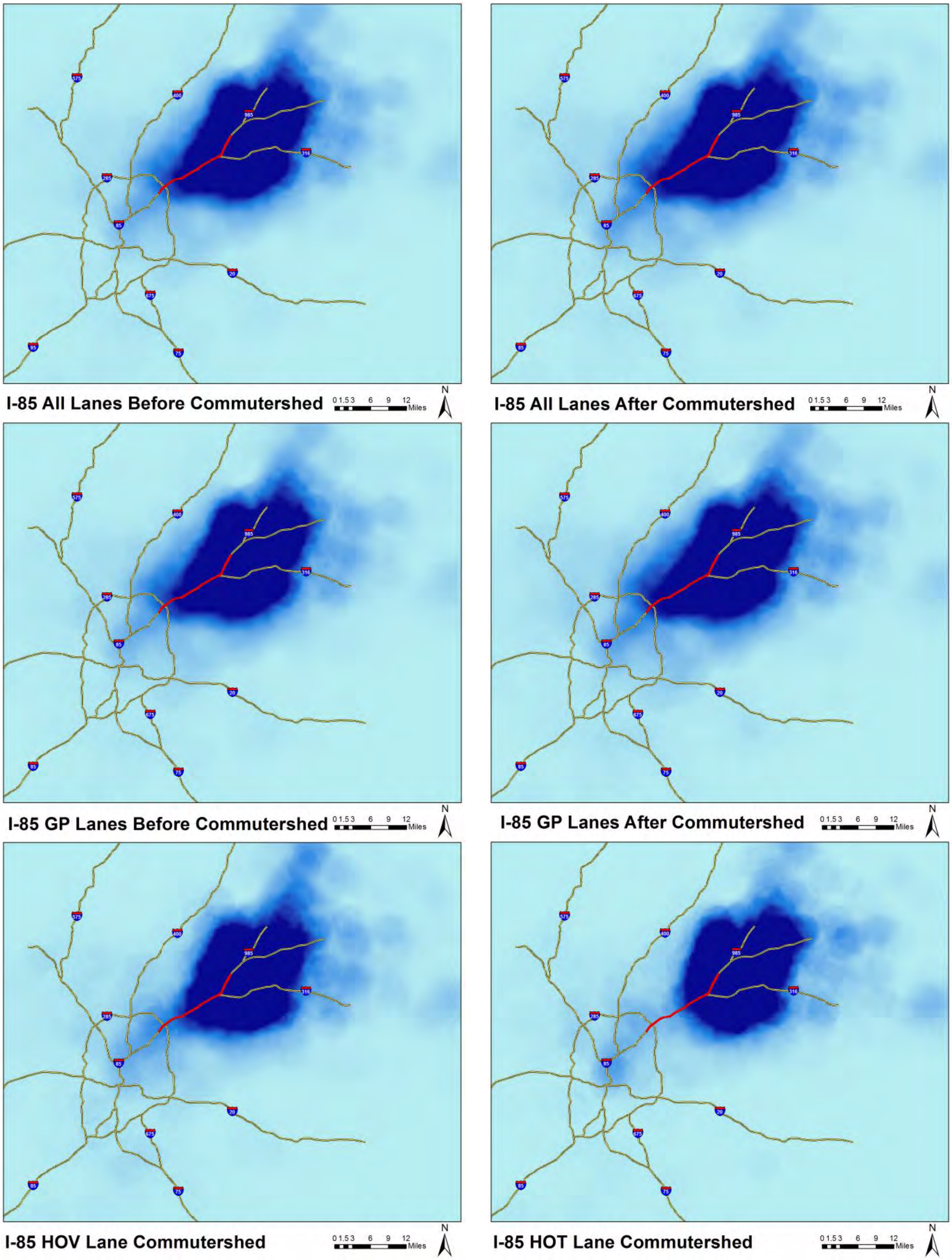


Figure 91 Fuzzy Membership Commutershed Maps

The same fuzzy commutersheds have been also developed for AM and PM peak period separately and illustrated in APPENDIX C, Table 56. During the AM peak period, the license plate data were collected in the southbound direction (toward the CBD), while they were collected along the northbound (toward suburban residential areas) during the PM peak period. Accordingly, it is expected that the majority of the morning trips are “home to work” trips, while there could be significant amount of shopping, recreation, and other trips in the afternoon.

To compare cell Fuzzy values before and after the conversion, the raster calculator in ArcGIS has been used. The raster calculator generates a new raster layer after applying the desirable numerical function on the input layers cell values. In this case, the fuzzy values of the before conversion have been deducted from the fuzzy values of the after conversion and multiplying by 100. The reason for multiplying by 100 is to build scale of impact between -100 to 100; therefore, the cells experienced 100 value of change in fuzzy membership, did the highest possible positive change. Similarly, the cells experienced -100 value of change in fuzzy membership did the highest possible negative change. Obviously zero value of change implies no change in corridor usage.

Figure 92 illustrates the output raster from subtracting the before conversion Fuzzy value from the after conversion Fuzzy values along all the lanes (multiplied by 100). Negative cell value (red cells) implies decrease in the corridor use, and positive cell value (green cells) implies increase in the corridor use (regardless of the lane).

The corresponding histogram of the cell values is illustrated in Figure 93. The cells with zero value have been removed from the histogram for visualization purpose. The output raster cell values range from -11 to +5 (compared to -100 to +100) which show very mild change in the corridor commutershed. About 17% of the raster cells (with respect to 95% ellipse) were positively impacted with 1.4 average increases in fuzzy value. However, 4% of cells were negatively impacted with a -3.6 average decreases in fuzzy value. Overall, the corridor commutershed expanded 13% in area (assuming commutershed to be all the cells with more than zero fuzzy value).

The main area that was negatively impacted is centered near the Lawrenceville, GA residential areas. While, this negative impact exists in both the AM and PM peak period (Figure 112, Figure 113), the positive impact mainly occurred during PM peak.

The average block group level income of the block groups intersected with green cells ($71,366) is only $906 higher than the average blockgroup level income of the red cells ($70,460). Considering that the average block group level income of the whole commutershed area is $65,730, income might not be the dominant reason for the observed reduction in corridor use.

Other than income, the presence of Lawrenceville Hwy parallel to the I-85 that connects this area to I-285 might be the dominant reason for observed decrease in corridor use Figure 94. The latter assumption makes more sense, considering the small difference between the average incomes of the positively and negatively impacted areas. Unfortunately, a sufficient number of before-after household travel diary surveys and panel studies were not conducted before and after HOT implementation to assess diversion of travel onto alternative routes.

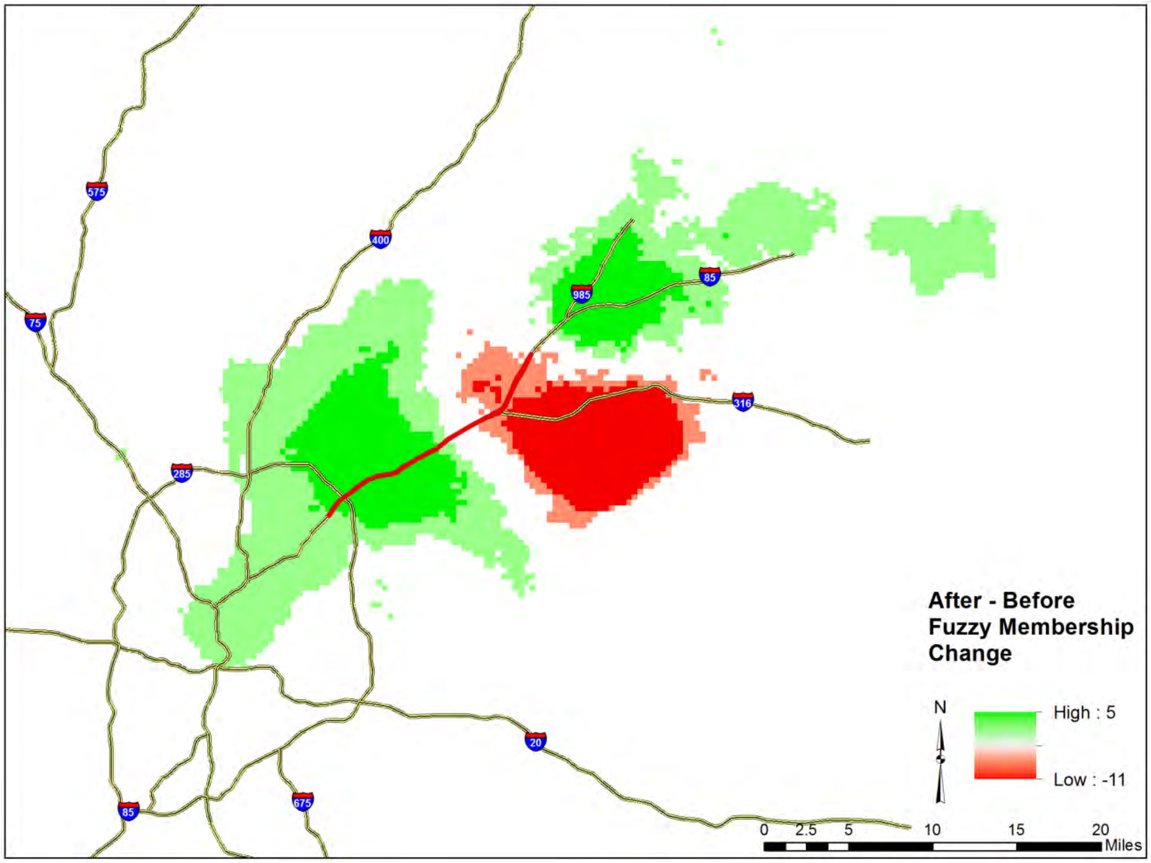


Figure 92 I-85 Before and After Conversion Commutershed Change

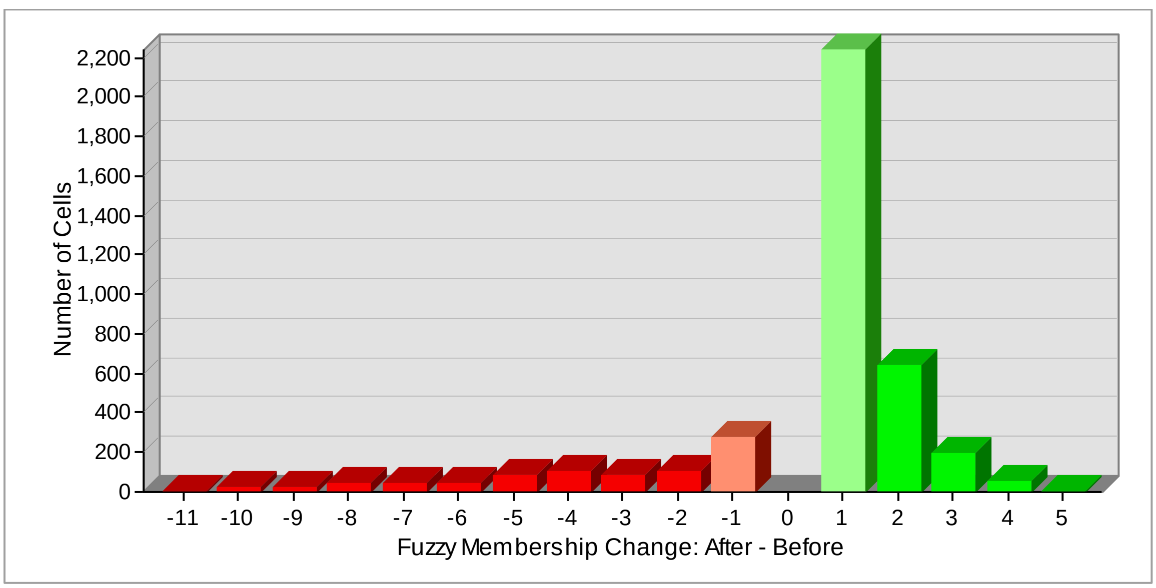


Figure 93 I-85 Before and After Conversion Commutershed Change Histogram (No change column is removed)

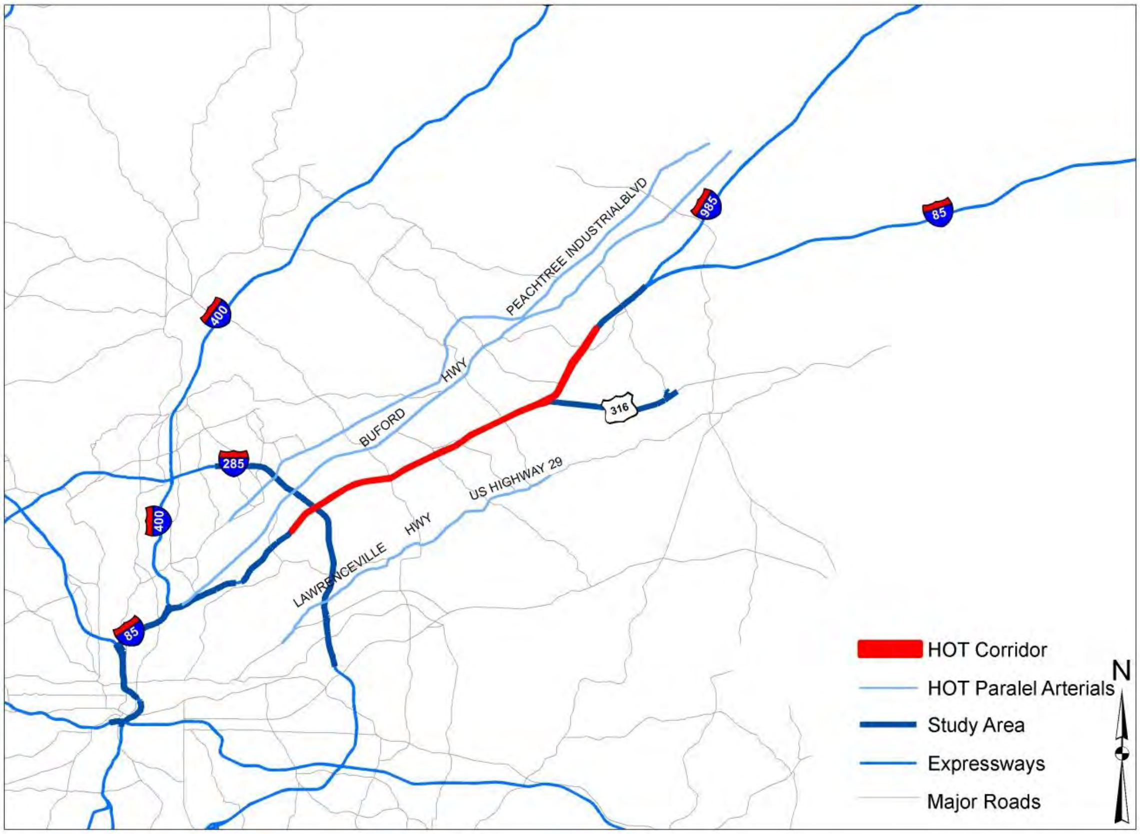


Figure 94 Parallel Corridors along I-85 HOT Corridor

The other potential reason might be switch to transit, considering the very close distance of the negatively impacted area to park and ride facilities (Figure 95). However, there are two issues related to this potential reason. One is that the Express bus ridership did not increase, even though bus service was increased by more than 18% (Guensler, et al., 2013). In this case, if there was a potential increase in transit ridership from Lawrenceville area, a similar decrease in transit ridership would have had to occur from other areas to yield a net zero change in transit ridership. However, there are other areas far north of the corridor which are also close to these park and ride facilities but did not experience the similar reduction in corridor use. Hence, diversion to transit does not appear to be a likely explanation.

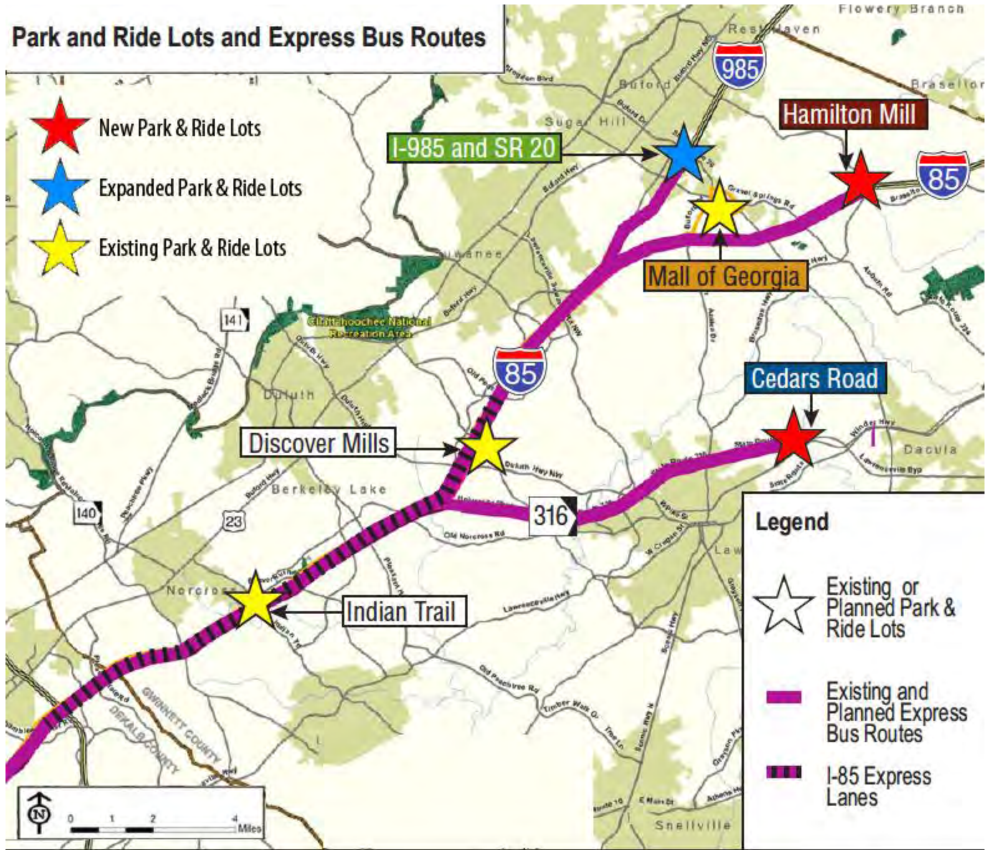


Figure 95 Park and Ride Facilities along I-85 HOT Corridor (Source: GDOT)

Similarly, Figure 96 shows the comparison between general purpose lanes before and after the conversion and Figure 97 shows the corresponding histogram. The change in the general purpose lanes and all the lanes commutersheds are very similar in terms of values and impacted areas. The output raster cell values range from -9 to +7 (compared to -100 to +100) which show very mild change in the corridor commutershed. And, 24% of cells (with respect to 95% ellipse) were positively impacted with an average of 1.8 in fuzzy value. However, 5% of cells were negatively impacted with a -2.6 average decrease in fuzzy value. Overall, the corridor commutershed along the general purpose lanes expanded 20% in area (assuming commutershed as all the cells with more than zero fuzzy value).

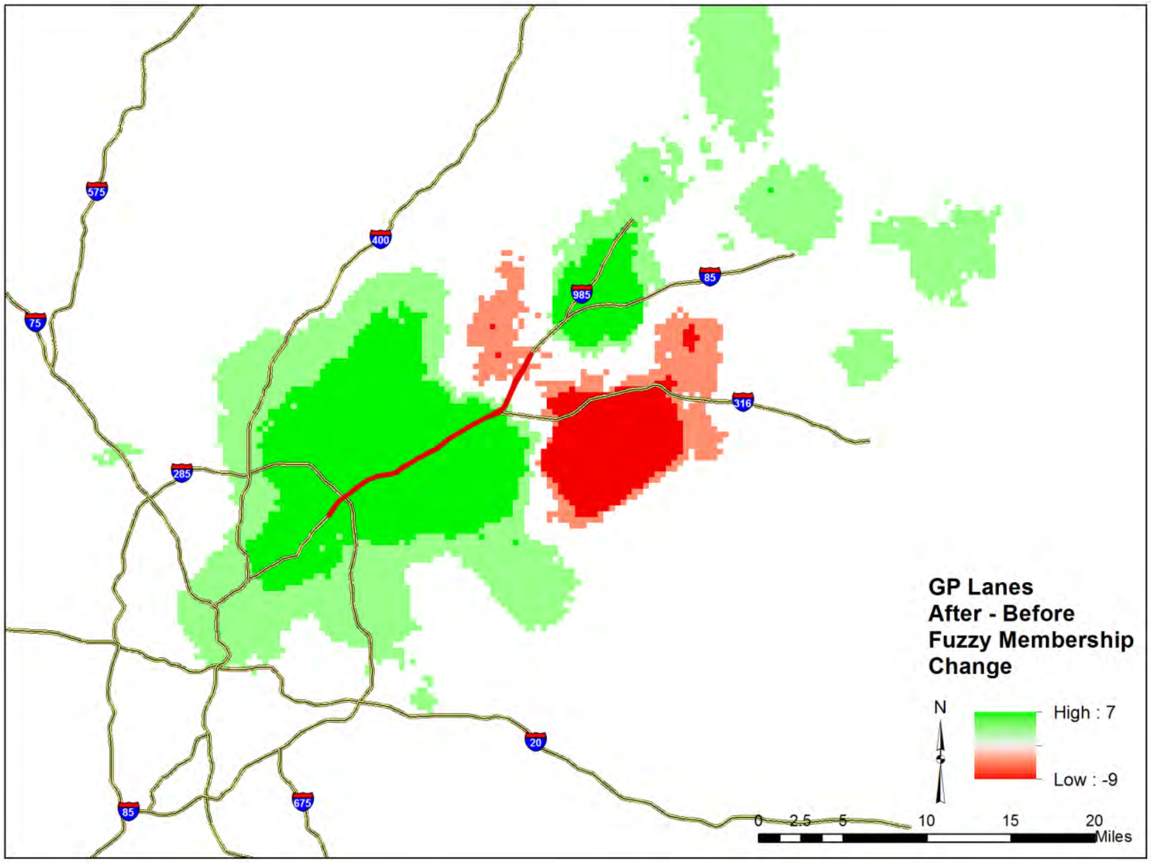


Figure 96 I-85 Before and After Conversion General Purpose Lanes Commutershed Change

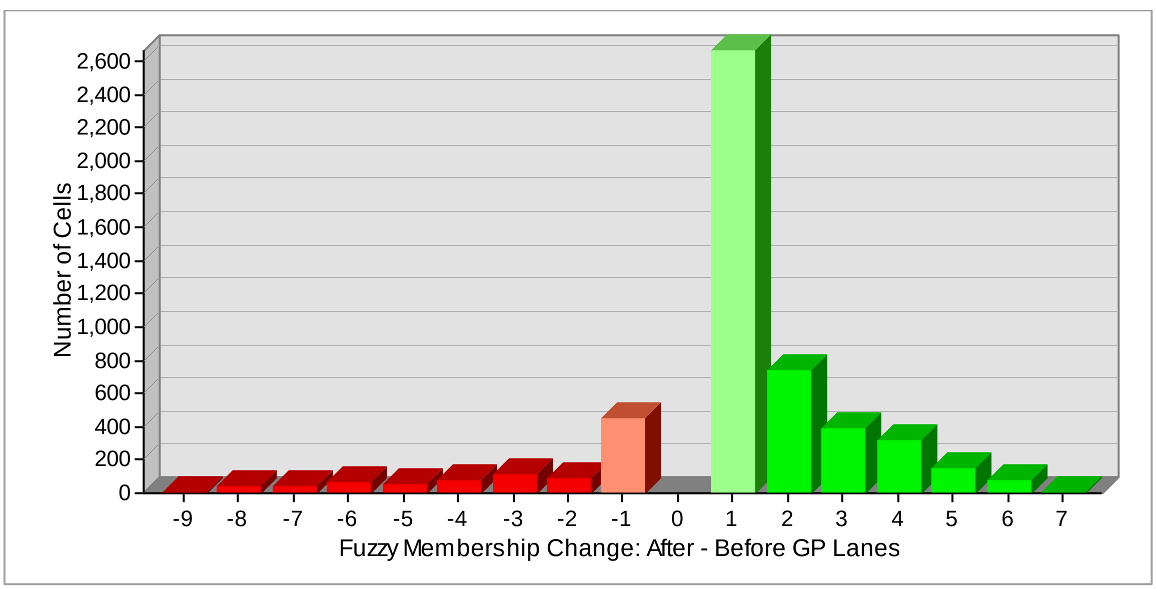


Figure 97 I-85 Before and After Conversion General Purpose Lanes Commutershed Change Histogram (No change column is removed)

Comparing the change in commutersheds, along general purpose lanes, between AM and PM peak periods illustrate that the majority of the positive change have only been occurred during PM peak period, whereas the negative change is existing in both AM and PM peak period (Figure 114, Figure 115). The negatively impacted Lawrenceville area persists during both AM and PM peak periods. However, some additional areas along Hwy 316 and north east part of I-85 have also been detected as negatively impacted areas.

The closeness to park and ride facilities as well as GA-400, and Peachtree Industrial Blvd could potentially describe part of the decreases in corridor usage. Other potential reasons might be change in time of commute to off-peak periods. A targeted household travel survey, or instrumented fleet, would provide the data necessary to assess the underlying reasons for the observed reduction in corridor use (but neither was conducted for the system).

The comparison between HOV lane and HOT lane commutersheds is illustrated in Figure 98, and the corresponding histogram is presented in Figure 99. Compared to what has been observed previously, the extent of the observed change along the managed lane is larger and significantly different. The cell values range from -33 to +20. 11% of cells (with respect to 95% ellipse) were positively impacted with 4.2 average increases in fuzzy value. However, 28% of cells were negatively impacted with -4.6 average decreases in fuzzy value. Overall, the corridor commutershed along the managed lanes retracted 18% in area (assuming commutershed as all the cells with more than zero fuzzy value).

The impact of the observed change in managed lane commutershed is almost opposite to the observed change in general purpose lanes commutershed, except that Lawrenceville areas that were negatively impacted along general purpose lanes were also negatively impacted along HOT lane (again indicating that overall traffic from this area may have diverted to Lawrenceville Highway). The positively and negatively impacted areas are quite similar during AM and PM peak period, except that the negatively impacted areas are more extended toward the Atlanta CBD during PM peak hour (Figure 116, Figure 117).

Contrary to what was observed for the general purpose lanes, the areas that were negatively impacted in HOT lane use are primary lower income Hispanic communities (mainly apartment complexes) along the I-85 corridor. Similarly, some areas in far northeast of Atlanta (outside of the metro area), which have lower income, experienced some negative impact as well. However, the majority of the negatively impacted areas along the HOT lane are positively impacted along the general purpose lanes; this implies that they may have switched to general purpose lanes.

The spatial distribution of the negatively impacted areas along HOT lane is very well correlated with distribution of low income areas (comparing Figure 15 and Figure 98). The average block group level income of the block groups intersected with green cells ($83,654) is $26,660 higher than the average blockgroup level income of the red cells ($56,985). Considering the average block group level income of the whole commutershed area which is $65,730, low income can strongly be considered as a potential reason for reduction in managed lane usage after the conversion.

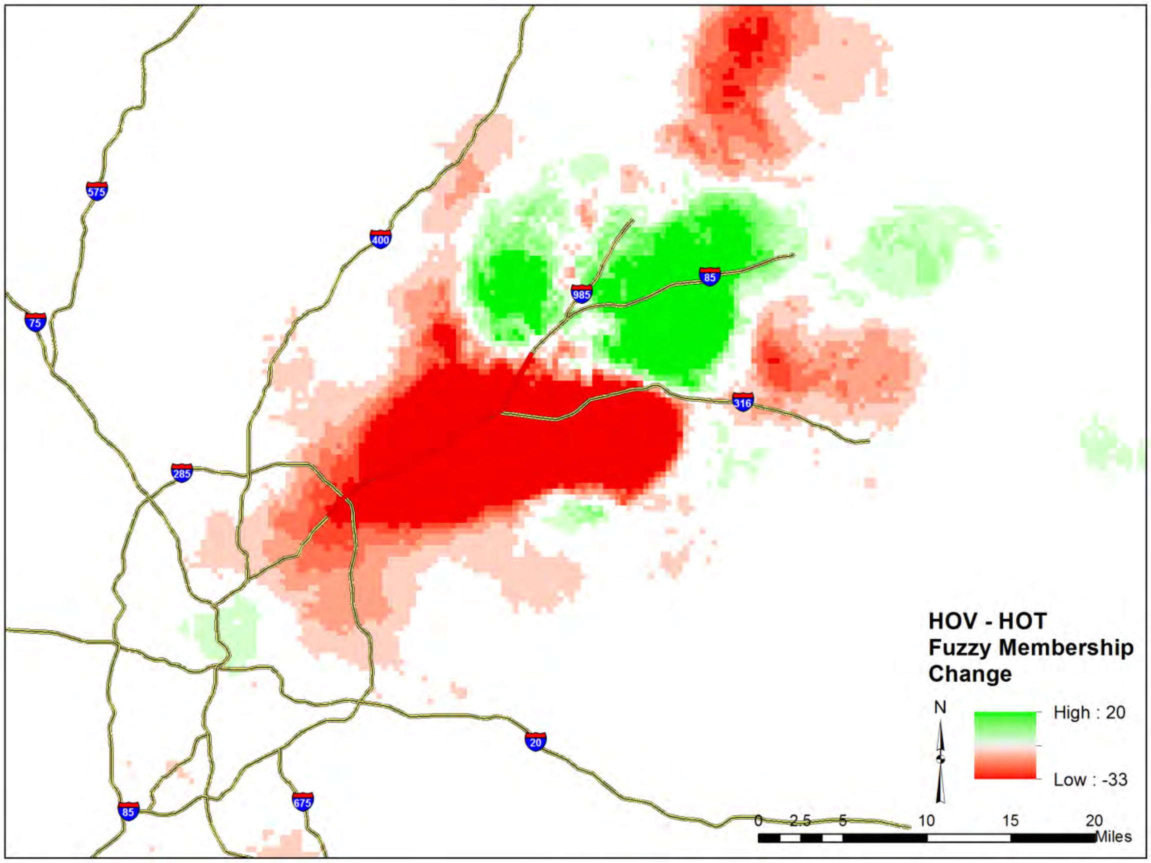


Figure 98 I-85 Before and After Conversion Managed Lanes Commutershed Change

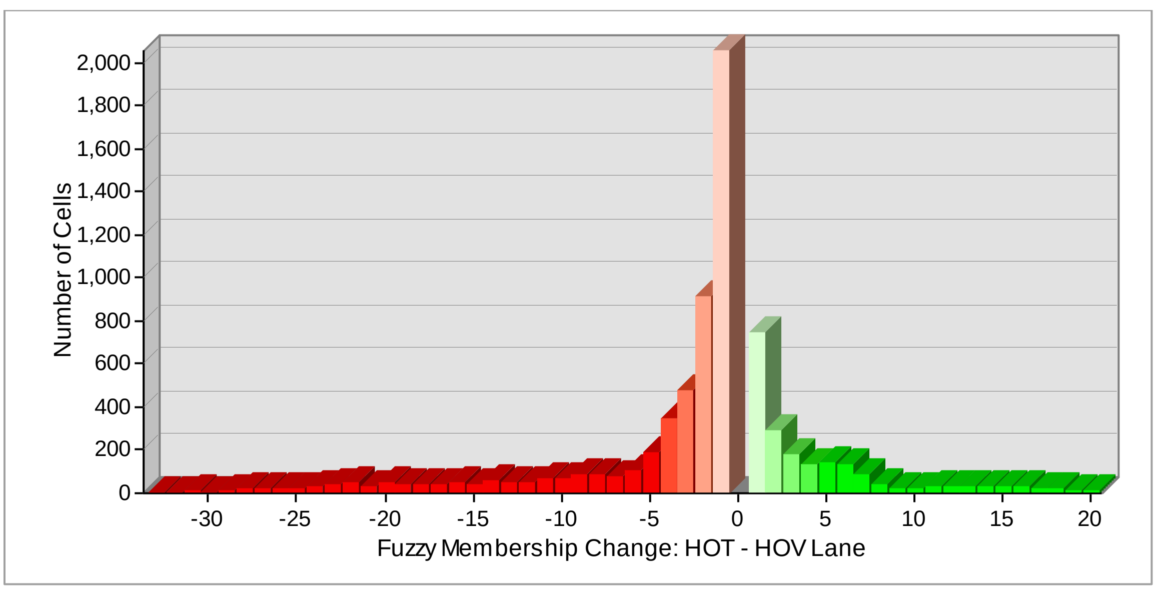


Figure 99 I-85 Before and After Conversion Managed Lanes Commutershed Change Histogram (No change column is removed)

Furthermore, Figure 100 which illustrates the spatial distribution of Hispanic population, shows a high spatial correlation with negatively impacted areas along the HOT lane. These findings correspond with household level statistical analysis results (Chapter 7), illustrating income and ethnicity are important predictors in managed lane usage. The positively impacted areas are more toward northeast, causing the observed shift in the center of the HOT lane commutershed toward suburban areas.

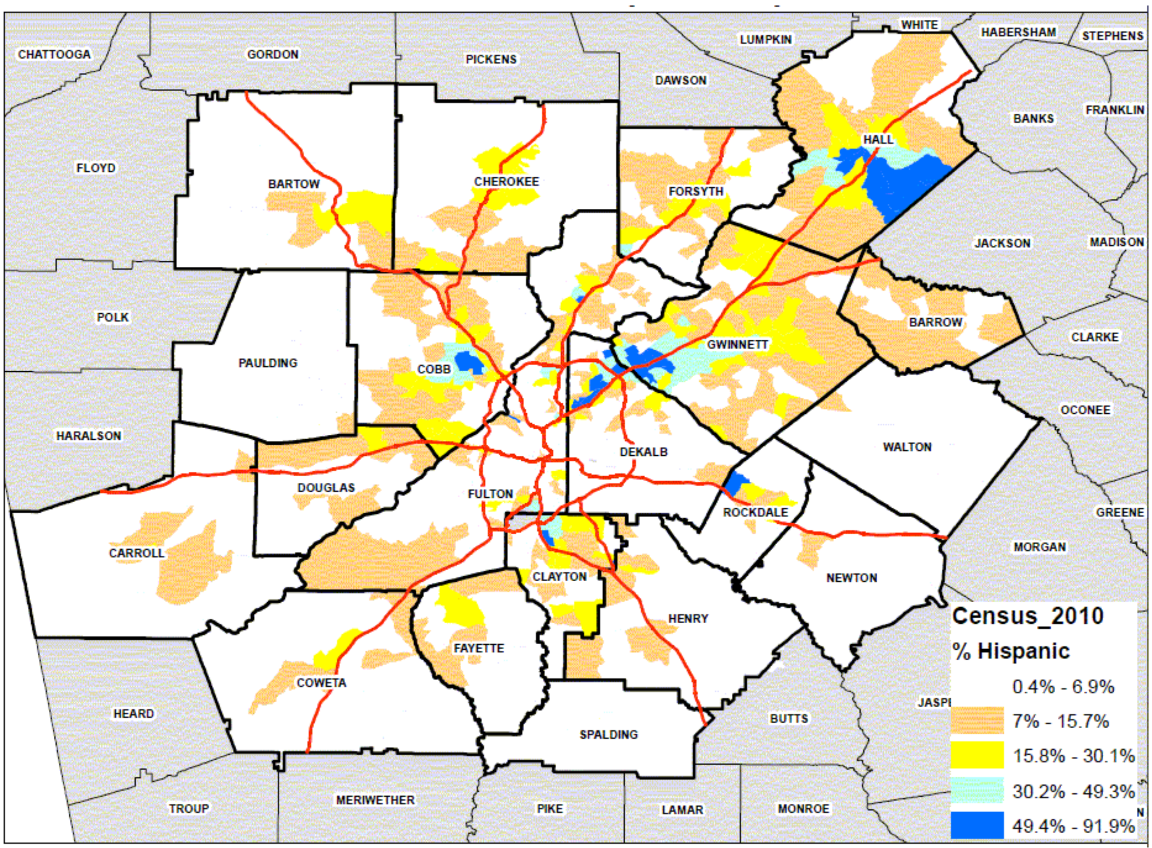


Figure 100 2010 Metro Atlanta Hispanic Population (source:census.org)

8.1.4. Discussion

Table 31 summarizes the Fuzzy values metrics for the developed raster layers. In summary, the managed lane commutershed area retracted 18% (339 sq. mi); the general purpose lanes commutershed area expanded 20% (340 sq. mi); consequently, all lanes commutershed area expanded 13% (224 sq. mi). Furthermore, the HOV lane commutershed is 112% larger than its adjacent general purpose lanes' commutershed. However, the HOT lane commutershed is 77% smaller than its adjacent general purpose lane's commutershed. The outcomes of both directional distribution and Fuzzy membership methods complement each other quite well.

Table 31 Commutersheds Fuzzy Membership Metrics

		Managed Lane	General Purpose Lanes	All Lanes
After Commutershed	# of Cells	13,840	17,924	17,095
	Area (sq. mi)	1,578	2,044	1,949
	Average Fuzzy Values	10.1	9.7	9.8
Before Commutershed	# of Cells	16,811	14,945	15,127
	Area (sq. mi)	1,917	1,704	1,725
	Average Fuzzy Values	9.5	10.8	10.6
Positively Impacted Cells	# of Cells	1,976	4,321	3,122
	Percent Area of 95% Ellipse	11%	24%	17%
	Area (sq. mi)	225	493	356
	Average Increase	4.2	1.8	1.4
Negatively Impacted Cells	# of Cells	5,048	884	806
	Percent Area of 95% Ellipse	28%	5%	4%
	Area (sq. mi)	576	101	92
	Average Decrease	-4.6	-2.6	-3.6
Total Change	Area	-18%	20%	13%

The overall expansion in the commutershed would not be observed if all the expansion in general purpose lanes commutershed was only the result of users switching out of the HOV lane and into the general purpose lanes. Users living farther away are using the general purpose lanes, despite the presence of the more reliable commute option. These long-distance users might have been using the corridor off-peak hour or might have using the alternative routes before the conversion. The increased demand on general purpose lanes could potentially explain the reason for slight decrease in general purpose lanes travel speed after the conversion.

However, it is difficult to assess this change without a detailed before and after survey data. It is important to note that in the existing Volpe survey, the after conversion sample was taken from the before conversion group. Therefore the Volpe survey overlooked the fact that the corridor users, after the conversion, may be different from before the conversion. Future before and after travel surveys should take into account the potential change in corridor users.

In a parallel study, February-April 2011 vehicle throughput data were compared with the same months in 2012, well after the HOT lane became operational on October 1, 2011. Vehicle throughput on the I-85 HOT corridor decreased by about 6.6% during the morning peak period, but only by about 2.8% during the afternoon peak period (Guensler, et al., 2013).

Furthermore, changes in traffic volumes at five control sites have been examined to see whether the noted changes on the HOT corridor were in line with changes in other locations. The control stations did not show any particular direction of change in traffic demand. One of the stations showed an increase in demand beyond 5% while another showed a decrease beyond 5%. The rest varied within a 5% bound. Therefore, although the commutershed has expanded, the corridor vehicle throughput has decreased. Therefore, expansion in commutershed doesn't necessarily mean increase in number of trips; because there are fewer trips from farther distance area compared to more closed areas.

With respect to the carpoolers, the percentage of 2-person carpools using the managed lane declined significantly, after the HOV2+ to HOT3+ conversion. Two-person carpools would have had to find a third commuter to operate on the HOT lane for free, or pay a toll to continue operating on the HOT lane. The percentage of 2-person carpools increased in all of the general purpose lanes, indicating that a significant number of carpools migrated to general purpose lanes and a significant fraction may also have disbanded. Therefore, the carpoolers' switch to general purpose lanes and disband in carpooling might explain the potential reason for the increase in general purpose lanes usage especially during PM peak period (Guensler, et al., 2013).

.Whereas the most probable reason for reduction in HOT lane usage is income and ethnicity, various potential reasons for reduction in general purpose lanes and consequently the corridor usage can be considered; these potential reasons include, but are not limited to, switch to alternative parallel corridors (Lawrenceville Hwy, Buford Hwy, Peachtree Industrial Blvd, and GA 400), switch to transit (considering the existence, expansion, and construction of park and ride facilities as well as increase in bus service), and switch to off-peak hour. To determine the dominant reasons for the observed impacts, an instrumented fleet study or wide-spread household travel survey is required.

In summary, all the different methods of commutershed analysis illustrate a substantially larger commutershed for HOV lane compared to the HOT lane. Therefore, any future studies including travel surveys, travel demand models and tolling and revenue studies should assign more weight on the central points of the commutershed to be more accurate in modeling or recruiting the HOT lane demand.

8.2. Spatial Autocorrelation

Spatial autocorrelation is a measure of the degree to which a set of spatial features and their associated data values tend to be clustered together in space (positive spatial autocorrelation) or dispersed (negative spatial autocorrelation). Spatial autocorrelation is initially defined by “Waldo Tobler” as the first law of the geography: “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1970).

Intuitively, one would expect that corridor usage increases as the blockgroups become closer to the corridor and decreases as the blockgroups become farther from the corridor. Therefore, the presence of spatial autocorrelation is expected across the blockgroups with respect to corridor usage. Accordingly, Figure 101, and Figure 102 illustrate the HOV lane and HOT lane usage normalized by number of block groups drivers (more than 16 years old). The reason for normalization is to prevent the spatial size of the block groups from impacting the results.

Managed lane usage by resident of the block groups closer together is spatially correlated. Measuring the strength of spatial autocorrelation before and after the conversion across blockgroups with respect to general purpose as well as managed lanes would illustrate how much the location is potentially impacting the users travel behavior. Meanwhile, it is important to consider that blockgroups closer to each other are more similar in terms of socioeconomic attributes as well.

8.2.1. Moran’s I

Moran's I is a measure of spatial autocorrelation developed by Patrick Alfred Pierce Moran (Moran, 1950). Moran’s I is defined in Equation 7, where N is the number of units indexed by i and j; X is the variable of interest; and w_{ij} is an element of a matrix of spatial weights. The spatial weight matrix in this study has been calculated based on the inverse function of distance.

$$\boldsymbol{I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} \, (X_i - \bar{X})(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2}} \qquad \textbf{Equation 7}$$

The value of Moran's I ranges between zero and one, where a value of zero indicates no spatial autocorrelation at all, while a value of one indicates 100% spatial autocorrelation. Table 32 summarizes the Moran's I calculation results. Basically, the general purposes lanes usage is more spatially correlated than is managed lane usage. With respect to the managed lanes, HOV lane usage is more spatially correlated than is HOT lane usage. Therefore, spatial location may be less important for HOT lane usage compared to HOV lane and general purpose lanes. This finding is intuitive considering the pricing scheme of the HOT lane. Therefore, socioeconomic characteristics especially income plays an important role and spatial location is less important with respect to the HOT lane usage.

Table 32 Moran's I Indices

	HOV Lane	**HOT Lane**	**GP Lanes Before**	**GP Lanes After**
Moran's I	0.574	0.458	0.650	0.655
Significance	0.000	0.000	0.000	0.000

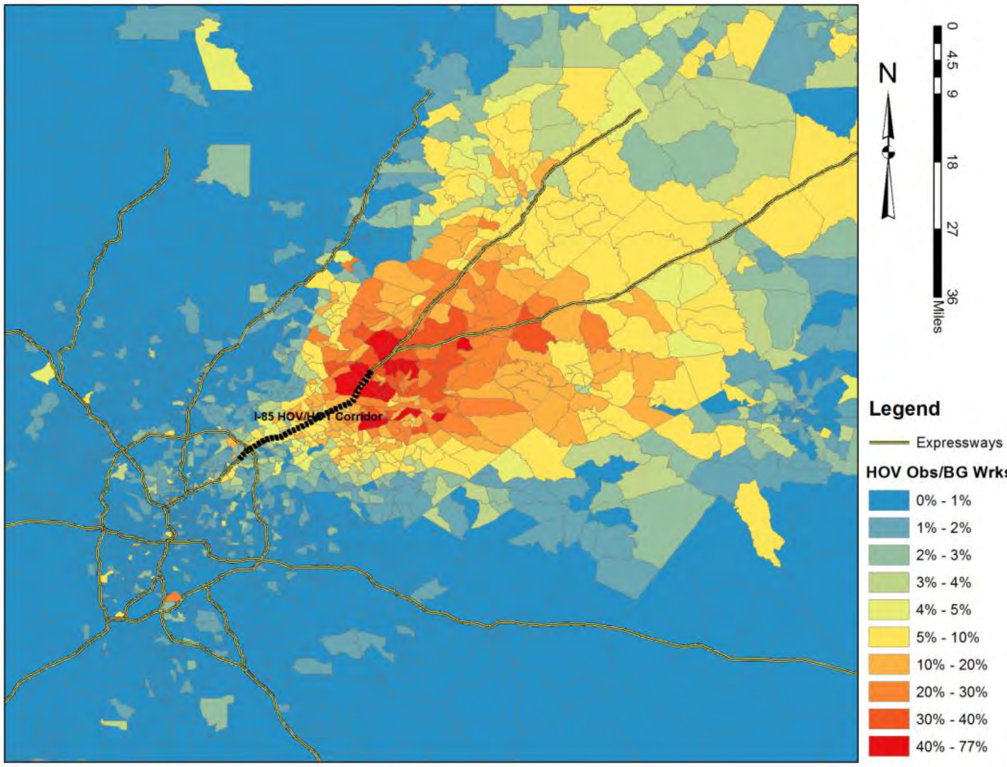


Figure 101 HOV Usage Thematic Map (BGs HOV observations normalized by number of workers)

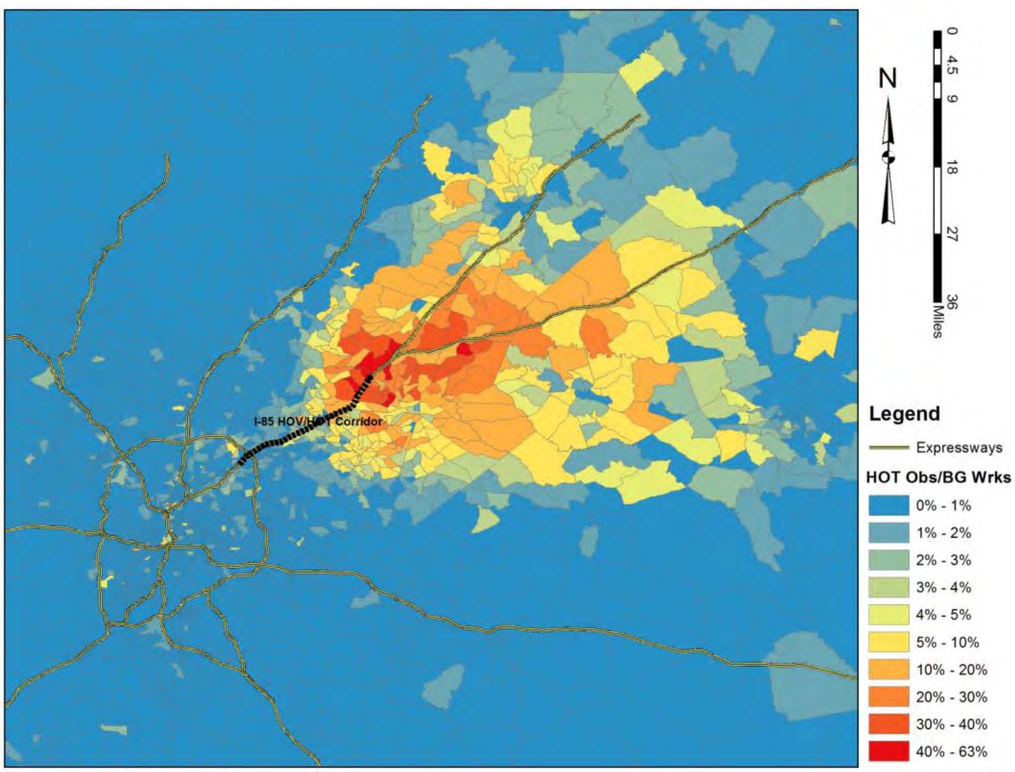


Figure 102 HOT Usage Thematic Map (BGs HOT observations normalized by number of workers)

8.2.2. Hot/Cold Spot Analysis

Now that corridor usage appears to be spatially correlated, it is interesting to examine the locations of Hot Spot and Cold Spot clusters. In ArcGIS, Hot Spot Analysis tool calculates the Getis-Ord Gi statistic for each feature in a weighted set of features. The Gi-statistic indicates whether features with high values or features with low values tend to cluster in a study area.

This tool works by examining each feature within the context of neighboring features. If a feature's value is high, and the values for all of its neighboring features are also high, it is a part of a hot spot. The local sum for a feature and its neighbors is compared proportionally to the sum of all features; when the local sum is much different than the expected local sum, and that difference is too large to be the result of random chance, a statistically significant Z score is the result. The resultant z-scores and p-values tell you where features with either high or low values cluster spatially.

The Gi statistic returned for each feature in the dataset is a z-score. For statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (hot spot). For statistically significant negative z-scores, the smaller the z-score is, the more intense the clustering of low values (cold spot) (Esri, 2013). Figure 103 illustrates the Hot/Cold Spot map for number HOV lane usage and Figure 104 illustrates the same map for HOT lane usage. The maps are color-coded by Giz-score standard deviation categories at 90%, 95% and 99% confidences. The area impacted by the HOV lane as Hot/Cold spots are substantially larger than HOT lane Hot/Cold spots. Specifically, the HOV lane Hot spot area is 142% larger than HOT lane Hot spot mainly toward northeast direction.

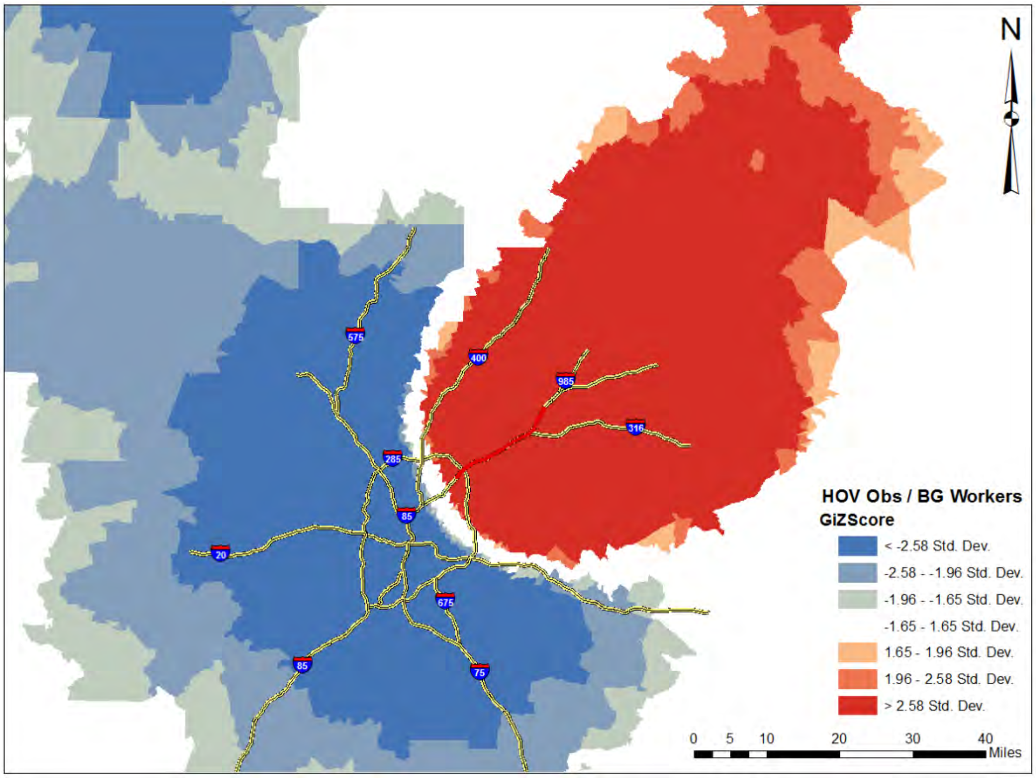


Figure 103 HOV Lane Usage Hot/Cold Spot Map

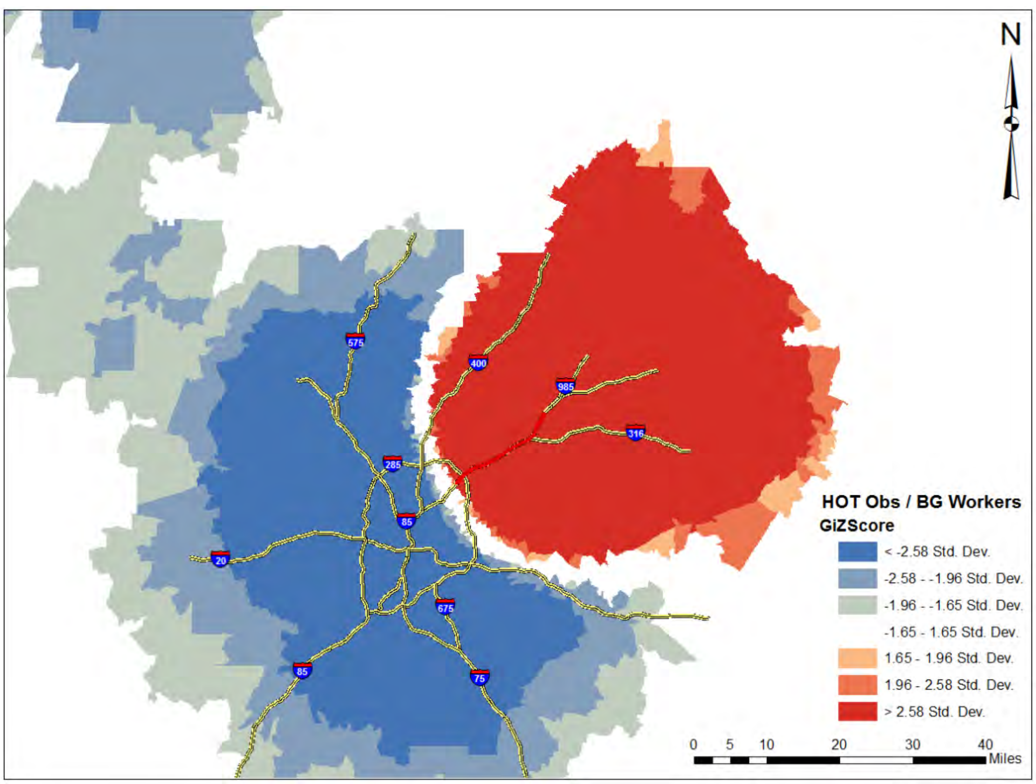


Figure 104 HOT Lane Usage Hot/Cold Spot Map

The extension of the Hot spot area of the HOV lane toward northeast, compared to the HOT lane Hot spot area, coordinates quite well with the directional distribution ellipse and block group level analyses results. In other words, the observed retraction in managed lane commutershed, after the conversion, has been again highlighted using the Hot and Cold Spot analysis. As it was discussed earlier, the HOV users who live in farther northeast area are not using the HOT lane as frequent as HOV lane. Meanwhile, the relative increase in the general purpose lanes use by the far-northeast users, once again, indicates the potential switch of the previous HOV carpoolers to general purpose lanes after the conversion.

8.3. Spatial Lag Model

The relationship between corridor usage and socioeconomic attributes has been studied in section 5.5 and 5.6 without considering spatial location at the household level. Furthermore, the block group managed lane usage has been modeled as a function of socioeconomic attributes and dummy variables for location (section 5.4.2). The latter model concluded that being among the block groups at far north-east of the corridor is significant for "HOV Usage" models while being among the block groups at near north-east of the corridor is significant for "HOT Usage" models. Therefore the demand for HOT lane may be higher from closer areas to the corridor compared to the HOV lane. These findings correlates very well with the commutershed analysis findings which shows the HOT lane commutershed is retracted compared to the HOV lane commutershed.

This last section of spatial analysis section is designed to assess whether incorporating the spatial correlation into the block group level model presented in Section

5.4 could improve the model goodness of fit. Different spatial weight matrices have been compared and the best model fit has been acquired by order 1 of queen contiguity. A queen weights matrix defines a location's neighbors as those with either a shared border or vertex (in contrast to a rook weights matrix, which only includes shared borders). Likelihood ratio tests for both HOV and HOT lanes usage diagnosed spatial dependence at 99% confidence, implying the need to incorporate spatial lag variable in the model. The same variables, which have been used in section 5.4.2, have also been used here to be able to compare the results. The dependent variables are HOV and HOT lanes usage. Logit transformation of the dependent variables has been used, because the dependent variables are between zero and one. The 2012 block groups in the I-85 commutershed are the units of analysis (N=2,012). Table 33 and Table 34 illustrate the models outputs. Both of the models' goodness of fit is very close and interestingly HOV model R^2 is slightly higher.

Using generalized linear models the HOT lane usage model ρ^2 is substantially higher (0.39) than the spatial lag model R^2 (0.18); however, the previously developed HOV lane model ρ^2 is very close to zero, compared to the spatial lag model R^2 (0.18). Other than the lag coefficient, which is statistically significant in both of the models, the number of significant variables for the HOV model is seven while HOT model has only four significant variables. Moreover, the direction of the income variables in the HOT model is opposite to the GLM model. In the GLM model, HOT usage is increasing up to household income of $134,967, while in the spatial lag model HOT lane usage is decreasing up to household income of $118,421. The other two significant variables in HOT lane use model are being female, which is positively correlated with HOT use, and drive to work which is negatively correlated with HOT use (similar to the GLM model output).

Table 33 HOT Lane Spatial Lag Model

Variable	Model 1	Model 2	Model 3(Final)
Lag coeff. (ρ)	0.417 (0.000)	0.416 (0.000)	0.417 (0.000)
Constant	-0.158 (0.677)	-0.145 (0.678)	-0.215 (0.415)
Household Size	-0.056 (0.372)	-0.055 (0.366)	
Income	-8.6e-006 (0.002)	-8.58e-006 (0.001)	-9.81e-006 (0.000)
Income Squared	3.8e-011 (0.003)	3.77e-011 (0.002)	3.86e-011 (0.000)
Age	0.003 (0.538)	0.003 (0.534)	
Vehicle Ownership	0.040 (0.480)	0.040 (0.472)	
High Education	-0.559 (0.129)	-0.566 (0.108)	
Female	1.051 (0.007)	1.054 (0.007)	1.061 (0.006)
Travel Time	0.000 (0.949)		
Drive to Work	-1.302 (0.000)	-1.314 (0.000)	-1.188 (0.000)
Work at Home	0.043 (0.942)		
R^2	0.184	0.184	0.182
LL	-3346	-3346	-3348
AIC	6716	6712	6709
S.E of regression	1.169	1.169	1.170

Table 34 HOV Lane Spatial Lag Model

Variable	Model 1	Model 2	Model 3(Final)
Lag coeff. (ρ)	0.367 (0.000)	0.366 (0.000)	0.368 (0.000)
Constant	-0.020 (0.939)	-0.061 (0.799)	-0.183 (0.400)
Household Size	0.137 (0.001)	0.137 (0.001)	0.144 (0.000)
Income	-8.60e-006 (0.000)	-8.52e-006 (0.000)	-1.03e-005 (0.000)
Income Squared	3.23e-011 (0.000)	3.20e-011 (0.000)	3.63e-011 (0.000)
Age	0.007 (0.032)	0.007 (0.035)	0.007 (0.040)
Vehicle Ownership	-0.061 (0.117)	-0.060 (0.125)	
High Education	-0.491 (0.055)	-0.492 (0.055)	
Female	-0.107 (0.696)		
Travel Time	-0.016 (0.000)	-0.016 (0.000)	-0.014 (0.000)
Drive to Work	-0.784 (0.000)	-0.796 (0.000)	-0.806 (0.000)
Work at Home	-0.758 (0.068)	-0.767 (0.064)	-0.819 (0.046)
R^2	0.199	0.199	0.197
LL	-2571	-2571	-2574
AIC	5167	5165	5166
S.E of regression	0.812	0.812	0.813

The direction of the HOV lane spatial lag model variables is more similar with the developed GLM model coefficients. More specifically, household size and age have positive correlation with HOV use while, income, vehicle ownership, high education, travel time, drive to work and work at home have negative correlation with HOV use.

In summary, generalized linear models, with dummy variables for location, work better for HOT lane modeling, whereas spatial lag models work better for HOV lane modeling at block group level. The underlying reason for this outcome is the existence of higher spatial autocorrelation across the block groups with respect to HOV lane use (carpoolers are living closer together) compared to HOT lane use. Therefore, the estimated coefficients of the final HOV model are explained in the following.

Because the dependent variable is Logit transformation of HOV usage, the odds ratios (exp(B)) of the coefficients are used for explanation. In terms of household size, the odds of block group HOV lane usage increases 15% (1-exp (0.144)) for one unit of increase in average block group household size. In terms of income, the odds of block group HOV lane usage decreases 10% (1-exp (-1.03e-005) ^10,000) for $10,000 increase in median block group income. However, this decrease trend is up to median household income of $141,873.

In terms of age, the odds of block group HOV lane usage increases 7% (1-exp (0.007) ^10) for 10 years increase in average block group age. In terms of commute travel time, the odds of block group HOV lane usage decreases 13% (1-exp (-0.014) ^10) for 10 minutes increase in average block group commute travel time. For 10% increase in percent of block group workers who drive to work, the odds of block group HOV usage increases by 8% (1- exp (0.806) ^0.1). Lastly, for 10% increase in percent of block group workers who work at home, the odds of block group HOV usage increases by 8% (1- exp (0.819)^0.1).

Comparing the odds ratios of two models estimating HOV usage (one with spatial lag term in this section and one without spatial lag term at section 5.4.2.), the odds ratios are larger for the model with spatial lag component. For example, the model without spatial lag component estimated the impact of increase in one unit of household size on the odds of block group HOV lane usage as 3%, whereas the same estimate with the model with spatial component is 15%.

Therefore, controlling for the spatial autocorrelation, the actual effects of socioeconomic attributes can potentially be estimated more accurately. However, the level of spatial autocorrelation (implemented in the model using spatial weight matrix) might not be the same in any future project at different locations. Hence, the application of a the spatially weighted models in this section for future projects should be under the assumption that similar spatial relationships exist between the block groups.

CHAPTER 9

MANAGED LANE SOCIO-SPATIAL MODELING FRAMEWORK

The objective of this chapter is to summarize the methods developed and applied in the previous dissertation chapters and to suggest a preliminary analytical framework that could be applied to future assessments of similar managed lane and tolling projects. The data collection methodologies, analyses, and results were illustrated in detail in the previous chapters. Based upon these results, this chapter proposes a step-wise framework for future socioeconomic analysis of transport systems. Similarly, traffic and revenue studies could use the resulting analytical framework to forecast the characteristics and probable travel behavior of target market in response to pricing.

The first step is to collect travel data of the current condition for use in forecasting future activity levels. Travel data specifically refers to elements that identify users of the corridor and their current travel behavior with respect to operations on the corridor. For example, in this study, license plate data were used to identify the households that were currently using the corridor before HOT conversion, as well as the associated frequency of use along HOV and general purpose lanes. Considering the available budget and desired accuracy of any future study, different methods of data collection and different amount of data could be collected.

Implementing license plate data collection (similar to what have been implemented in this study) would cost approximately ¢10 per complete license plate identification (i.e., plates that yield matched records in the registration database with fewer than eight registered vehicles per address, and the registration address is in reasonable proximity to the corridor). On average, one two-hour session of data collection on a six lane corridor (12 lane-hours) produces 7,719 complete license plate records at peak hour and costs only $768 (including the cost of manual license plate extraction). Accordingly, one lane hour collected video produces 643 correct license plates and costs about $64.

The estimated costs of travel data collection is based on the applied methodology and the desired amount of data. More advanced methodologies such as Automatic License Plate Readers (ALPR), RFID tag readers, and cell phone data could certainly decrease the cost of travel data collection (compared to travel diary data collection) and improve the efficiency and labor and time intensiveness in the future.

Based on the collected travel data, the next step is to establish the corridor commutershed. Accordingly, the commutershed could be developed at household level or at the block group level. The household level analysis needs enough data to provide a reliable frequency of corridor use per household as well as identifying the frequent users (top 5% frequent license plates in this study). In this study, each household was observed an average of five times during the 1860 lane-hours data collection. This large amount of data (1,196,433 complete license plates matched to 241,466 households) enabled this research to establish target market groups and support household-level models. Although large datasets enhance the accuracy of the results of future studies, collecting such a large amount of data is not necessarily required for all future studies.

Future studies can assess the amount of data that they could collect using license plate visual capture method (which has been applied in this study) and limit data collection within budget constraints. The numbers of collected households and (correct) license plates as a function of amount of data collection are illustrated in Figure 105. The provided functions can be used for future data collection estimation. Because there are so many regular users on a commuter corridor, plate data collection yields diminishing returns with respect to identification of new households. For example, if we assume that 1000 lane-hours yield 198,000 households, we can get 60% of them at 500 hours.

Figure 106 illustrates the cost of data collection considering ¢10 per correct license plate estimation. For the explained example, the cost of collecting about 198,000 households (or 653,000 license plates) is approximately $66,000. In the context of a $100 million project, this is insignificant. However, this is the cost of data collection for only one corridor (15.5 mile in this case). Data collection cost for projects that involve multiple corridors would be much higher.

Figure 107 illustrates the relationship between average observation frequency per household as a function of amount of data collected (lane hours) with the blue line. The estimated power functions can be used by future researchers to estimate average observation frequency per household. Whereas the average observation frequency is estimated across all the license plates, the minimum frequency of the frequent corridor commuters (top 5% frequent license plates) has also been illustrated with the red line. This latter variable is important for identifying the frequent users for the application of developed market-based models. The slope of variation for average observation frequency is relatively flatter than the minimum frequency of top 5% users. For example, by collecting 252 lane hours' worth of data (21 two-hours session for a six lane highway which corresponds to one season of data collection in this study), the average frequency is 2.3 and the minimum frequency of top corridor users is seven, which is large enough for applying all the developed models in this study.

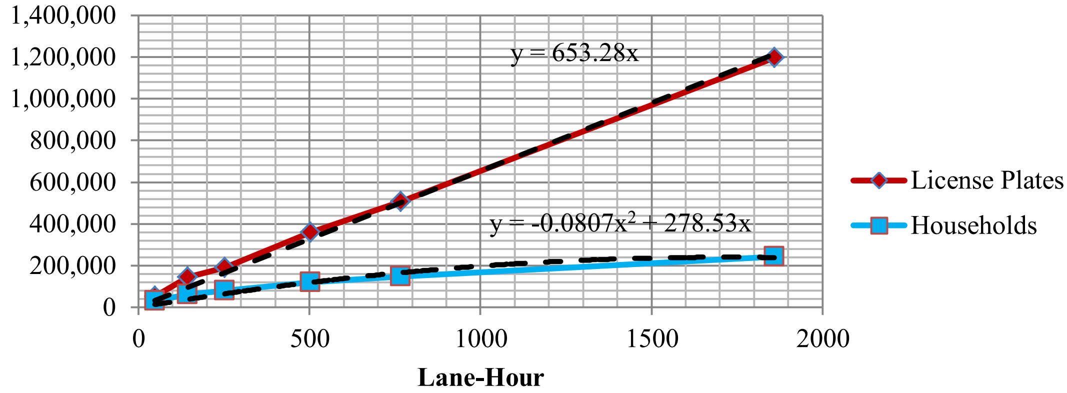


Figure 105 License Plates and Households Count as a Function of Amount of Collected License Plate Data

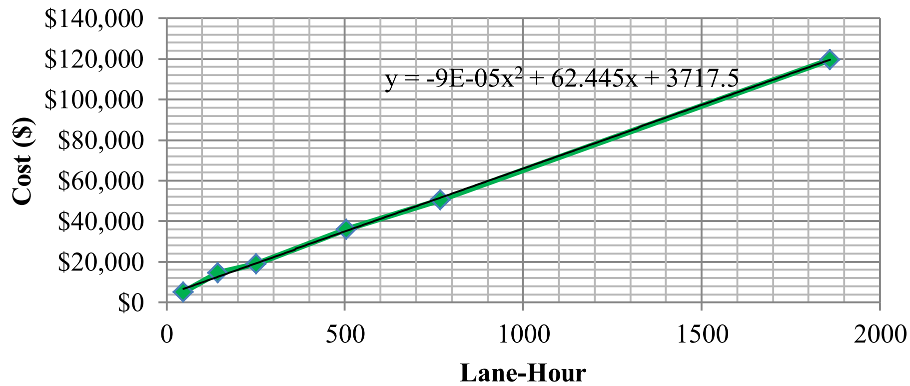


Figure 106 Data Collection and Processing Cost Estimate as a Function of Amount of Collected License Plate Data

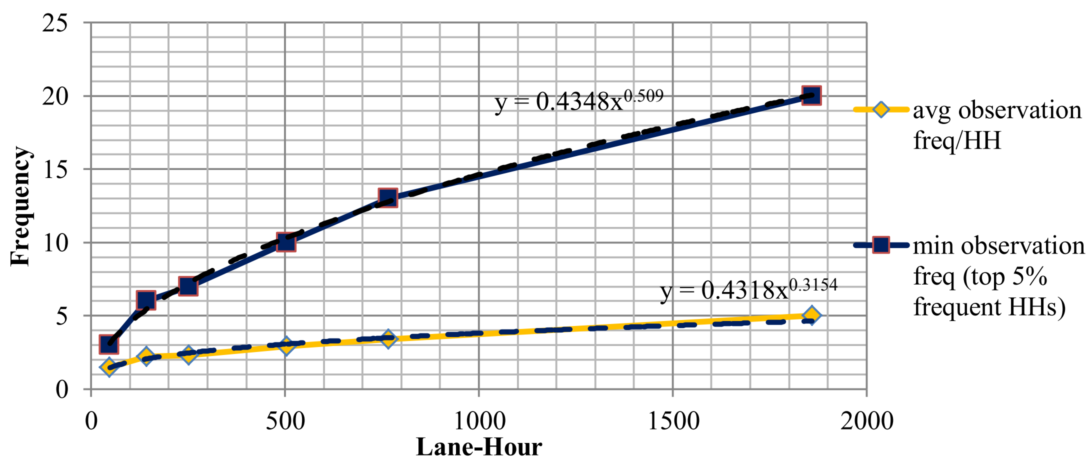


Figure 107 License Plate Data Frequency of Observation per Household as a Function of Amount of Collected License Plate Data

Lastly, the sensitivity analysis of this study shows that license plate data along a corridor are more sensitive to data collection location compared to time (a.m. vs. p.m.), and are more sensitive to data collection time compare to weekday. Therefore, it is better to spread the data collection locations across the corridor and make sure to collect at both morning and afternoon peak hour.

The third step is to acquire socioeconomic and demographic data. For block group level analysis, the most recent publicly available American Community Survey data should be used. The block groups that need to be incorporated in the analysis are the ones that intersect with the developed commutershed. However, as illustrated in Chapter 7, the modeling results based upon disaggregate household level data are preferable.

For household level models, a full set of marketing data, which has been used in this study, cost ¢10 per household. The socioeconomic data should be purchased for the addresses identified in the previous steps. Considering the multi-million cost of the entire project, the entire cost of license-plate data collection and socioeconomic data acquisition is perfectly reasonable.

The socioeconomic data can also be collected using other methods such as surveys and cell phone apps. The sample size of traditional surveys is very small compared to the large number of collected households and not strongly recommended for project-level analysis. However, the application of more innovative forms of surveys such as cellphone apps as part of a before-after panel survey may be justifiable. More specifically, if the travel data have been collected using cell phone data, the collection of socioeconomic data with cell phone apps would make the data collection process less expensive and more efficient.

In general, any big-data collection methodology, which can produce both travel and/or socioeconomic data, could be applied to the models developed in this study. Specifically, the main proposed methodology in this study is the concurrent application of license plate data, registration data, and marketing data.

Figure 108 illustrates the general steps an analyst should accomplish to implement the models developed in this study at any desired level of resolution. Once the socioeconomic data have been acquired for the corridor users, the last step is to run the developed models considering the socioeconomic and/or travel data as model input. The output of the models would be different elements of travel response to managed lanes. Table 35 illustrates eight different models that have been developed throughout the study and the associated specifications. Model 1 and Model 2 are at the block group level and Model 3, Model 4, Model 5, Model 6, and Model 8 are at the household level.

Model 1 predicts the HOT lane proportion of usage at each blockgroup using Generalized Linear Models (Probit). This model is based on the previous HOV lane usage as well as location dummy variables in addition to block group level socioeconomic attributes. The developed commutershed of the corridor would be used to select the block groups to be included in the modeling. Meanwhile, a version of the model has also been developed only based on block group socioeconomic attributes for the situations that data collection could not be conducted (to acquire HOV lane usage); however, the prediction power of the model decreases.

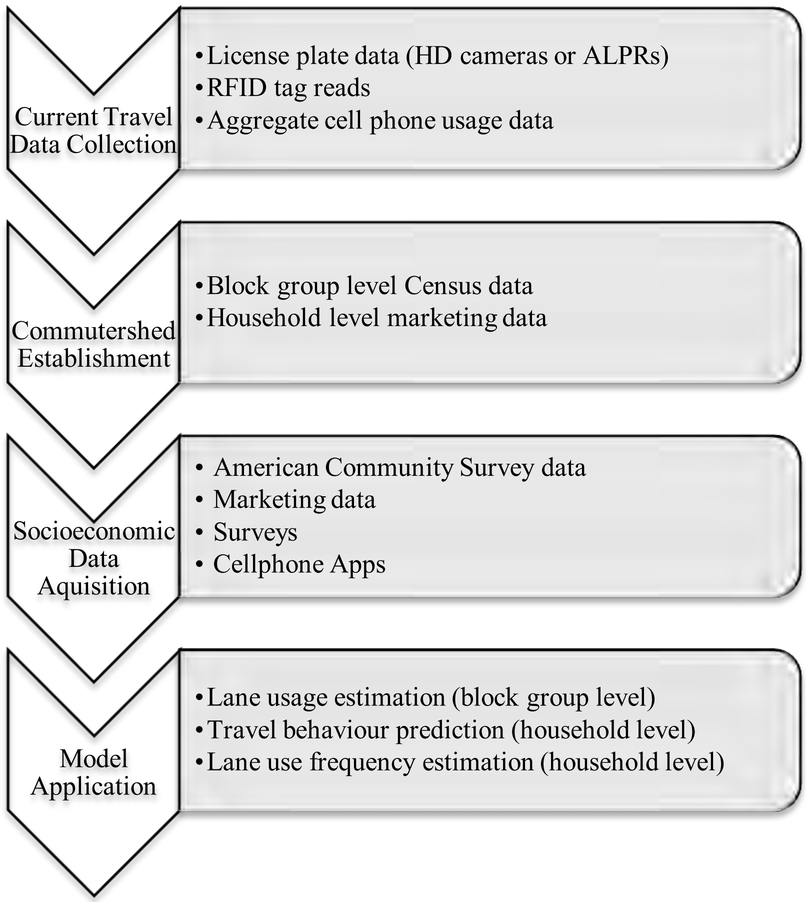


Figure 108 Modeling Framework for Socioeconomic Analysis of Managed Lanes

Model 2 predicts the HOV usage using spatial lag models. The presence of high spatial autocorrelation between the block groups with respect to the HOV lane usage makes this model fit better than Generalized Linear Model. To conduct this model, block group level socioeconomic data and a spatial weight matrix (can be easily developed using GIS soft-wares) are required.

The next two models were developed using the top 5% frequent corridor users (representing 42% of observations). Model 3 predicts the HOV frequent users response to pricing which is either continuing using HOT lane or switch to GP lanes. Similarly, model 4 predicts the GP frequent users response to a prices managed lane which is either continuing using GP lanes or switch to HOT lane.

Table 35 Developed Models for Socioeconomic Analysis of Managed Lanes

Model	Table	Predicted	Unit of Analysis	Predictor Data Source	Model Type	Prediction Power
1	10	HOT Usage	BG	ACS	GLM (Probit)	0.46
2	35	HOV Usage	BG	ACS	Spatial Lag	0.20
3	26	HOV to HOT/GP	HH	Marketing Data	Logistic Regression	0.05
4	27	GP to HOT/GP	HH	Marketing Data	Logistic Regression	0.32
5	29	HOV Use Frequency	HH	Marketing and Travel Data	Count Model	0.53
6	30	HOT Use Frequency	HH	Marketing and Travel Data	Count Model	0.68

While these models have been developed based on the specific behavioral response to the HOV2+ to HOT3+ conversion, they potentially might be applicable in other configuration as well. For example, Model 3 could be used in any project to predict travelers' response to a toll lane alternative. Continuous implementation of before-after data collection and model derivation will refine the models for specific situations and result in ongoing improvements to these methods.

Model 5 is a count model, which predicts the number of HOV lane use frequency as a function of socioeconomic attributes and total corridor use frequency for each household before the conversion. Similarly, model 6 is a count model, which predicts the number of HOT lane use frequency as a function of socioeconomic attributes and total corridor use frequency for each household after the conversion.

The last two models are probably the most valuable models in this study; because, the goodness of fit is the highest and also the model output which is managed lane use frequency is very valuable for use in traffic and revenue studies. The estimated managed lane use frequency can be imported to the toll and revenue estimation models and proper pricing and budget estimation can be conducted in the planning phase. For example, if tolling and revenue analysts apply the same model for another corridor and predict high potential managed lane demand (higher than lane capacity), they can predict that the I-85 pricing algorithm would not be appropriate and higher pricing scheme in general would be required. Although the last two models are probably the most valuable and accurate models in this study, the fact that they need household frequency of corridor use as an input variable makes these models expensive as well, given the need for large amounts of data.

To illustrate the application of the models developed in this study, Atlanta northeast project has been considered as a case study. Considering the budget and time constrains of this dissertation, it was not possible to collect license plate data. Therefore, a small size license plate data collected in 2006 has been used for commutershed development for illustration purposes.

9.1. Case Study: Atlanta Northeast Project (I-75/I-575)

9.1.1. Project Description[6]

The Northwest Corridor Project corridor extends northwest along I-75 from Akers Mill Road to Hickory Grove Road, and along I-575 from the I-75/I-575 interchange to Sixes Road. The project corridor is northwest of downtown Atlanta, Georgia. The study area encompasses the suburban cities of Marietta, Kennesaw, and Acworth and several unincorporated communities in Cobb and Cherokee counties. The area is home to a substantial share of the metro region's population as well as several business centers, two large regional shopping malls, Dobbins Air Force Base, and numerous major corporations.

The main freeway serving the Northwest Corridor is I-75. It is the primary route for commuters traveling to jobs within the region and to downtown Atlanta. As a major north-south route through Georgia, I-75 also serves the transportation needs for regional travel and freight trucking. In downtown Atlanta, I-75 is merged with I-85 through Downtown Atlanta. North of the I-75/I-85 split in Midtown Atlanta, I-75 turns to the northwest and intersects with the I-285 perimeter highway around Atlanta. Between I-85 and I-285, I-75 is ten lanes, with four general purpose lanes and an HOV lane in each direction. The existing HOV system ends at Akers Mill Road just south of I-285. North

---

[6] Draft Environmental Impact Statement for the Northeast Corridor Project; Prepared by GDOT and FHWA, May 2007.

of I-285, I-75 widens to as many as 15 lanes and then narrows again to eight then six lanes total north of the I-575 interchange.

I-575 extends northeasterly from its interchange with I-75 in Cobb County into Cherokee County. The I-575 corridor is experiencing rapid urbanization but currently has only has four general-purpose lanes, two in each direction.

The Northwest Corridor Project proposes to make transportation improvements to I-75 and I-575 in the Northwest Corridor. The proposed improvements include the addition of managed lanes on both-75 and I-575 within the project area from Akers Mill/I-285 north to Hickory Grove Road. Access points along I-75 are proposed at I-285, Terrell Mill Road, Roswell Road, I-575, Big Shanty Road, and Hickory Grove Road.

Along I-575, the proposed improvements include the addition of a managed lane system in the median between the I-75/I-575 interchange and Sixes Road. The managed lane system on I-575 could, as an option, include slip ramp access between the managed lane and the general purpose lane systems rather than direct access interchanges at Big Shanty Road, Shallowford Road, and Dupree Road. Figure 109 illustrates the project map.

Figure 109 Atlanta Northeast Corridor Project (Source: www.traffictechnologytoday.com)

9.1.2. Data

To build a 95% commutershed for I-75, the license plate data collected in 2006 study conducted by Jennifer Nelson, Dr. Guensler and their colleagues have been used (Nelson, et al., 2008). For one week during June 2006, researchers collected the license plate characters of passenger vehicles (i.e., cars, SUVs, pickup trucks, minivans, and conversion vans) observed traveling in the morning peak period direction on I-75 southbound (SB) at Windy Ridge Parkway. Data were collected for 1 to 1.5 hour (always overlapping the 7:30 to 8:30 a.m. period with the highest traffic volume) by using overpasses as observation points. From their overpass vantage points, researchers collected data on three general-purpose travel lanes at each location using spotting scopes, voice recorders, and video recorders. License plate data for trucks, buses, and out-of-state vehicles were not collected.

Approximately 5000 license plates were collected. The average frequency of observation per household is 1.0 which eliminates the opportunity to extract the main corridor commuters. Furthermore, no marketing data were purchased for this analysis. Therefore, the only option is to conduct block group level modeling for HOT lane usage (Model 1).

The total of 5000 license plates has been imported into ArcGIS and the 95% directional distribution ellipse has been developed (Figure 110). The block groups that intersect with the commutershed have been selected and required socioeconomic attributes have been extracted from American Community Survey. Using "Spatial Join" in ArcGIS, the license plates have been joined to block group layer and using the "frequency" tool, the numbers of observed frequency per block group have been appended to the block group layer. Hence, the block group layer DBF file has all the required socioeconomic attributes in addition to frequency of license plate observation per block group together.

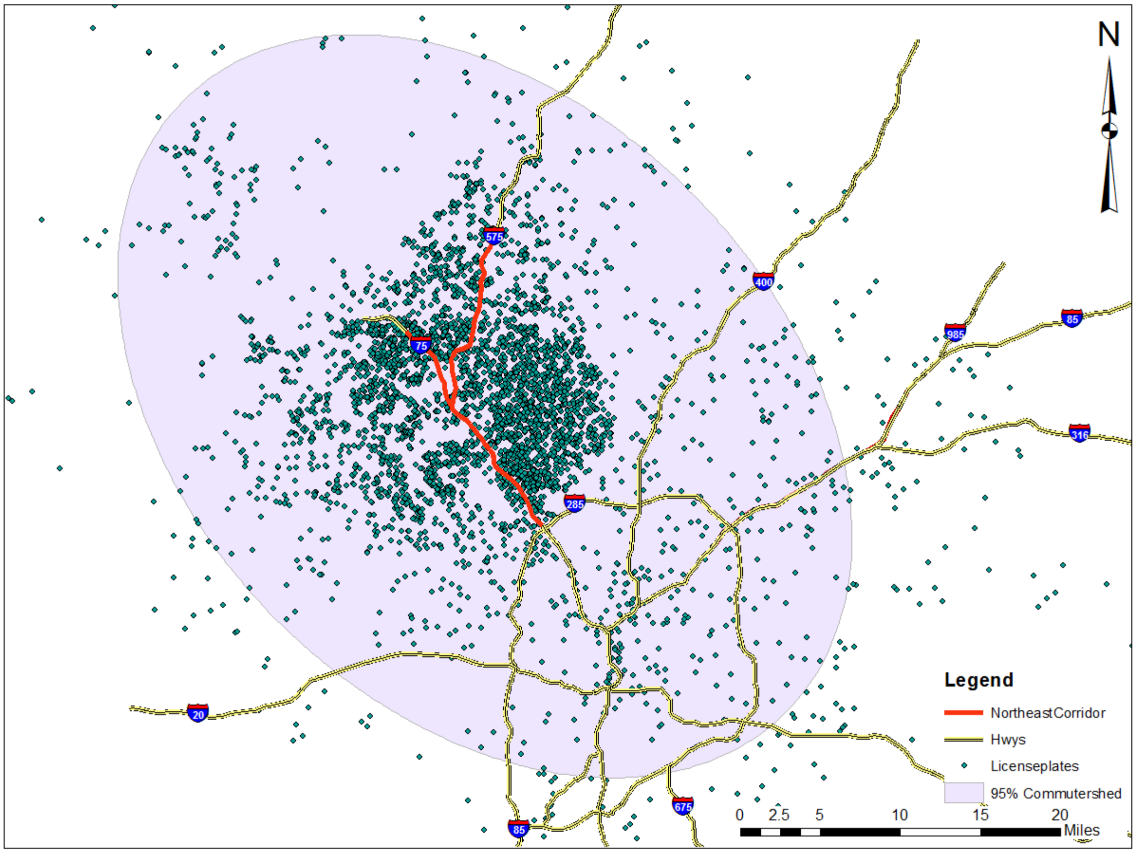


Figure 110 Northeast Corridor Commutershed Map

The next step is to import the block group layer DBF file to SPSS to apply the developed model. It worth mentioning, that all the developed models have been saved in XML format for future application. Using "Scoring Wizard" in SPSS, Model 1 has been applied on the data and the model outputs have been estimated. The model outputs include the frequency of managed lane usage per block group (with respect to the total frequency of corridor use) and the standard error of the predicted value. Dividing the predicted value by total frequency of observation produces the estimates for managed lane usage ratio at each block group. The average standard error of the predicted value is 0.01, and the standard deviation of the standard error is 0.001.

The estimated managed lane usage ratios have been illustrated in Figure 111 in five quintiles. As expected, the high income residential areas mainly located between the GA-400 and I-75 have been identified as potential high frequency managed lane users, whereas the low income residential areas along the corridor and far north-west of the corridor have been identifies as potential low frequency managed lane users. This finding corresponds with the observed managed lane usage along I-85 corridor (Figure 35). The visualization method in both of the maps is five quintile groups, although the actual categories are not equal.

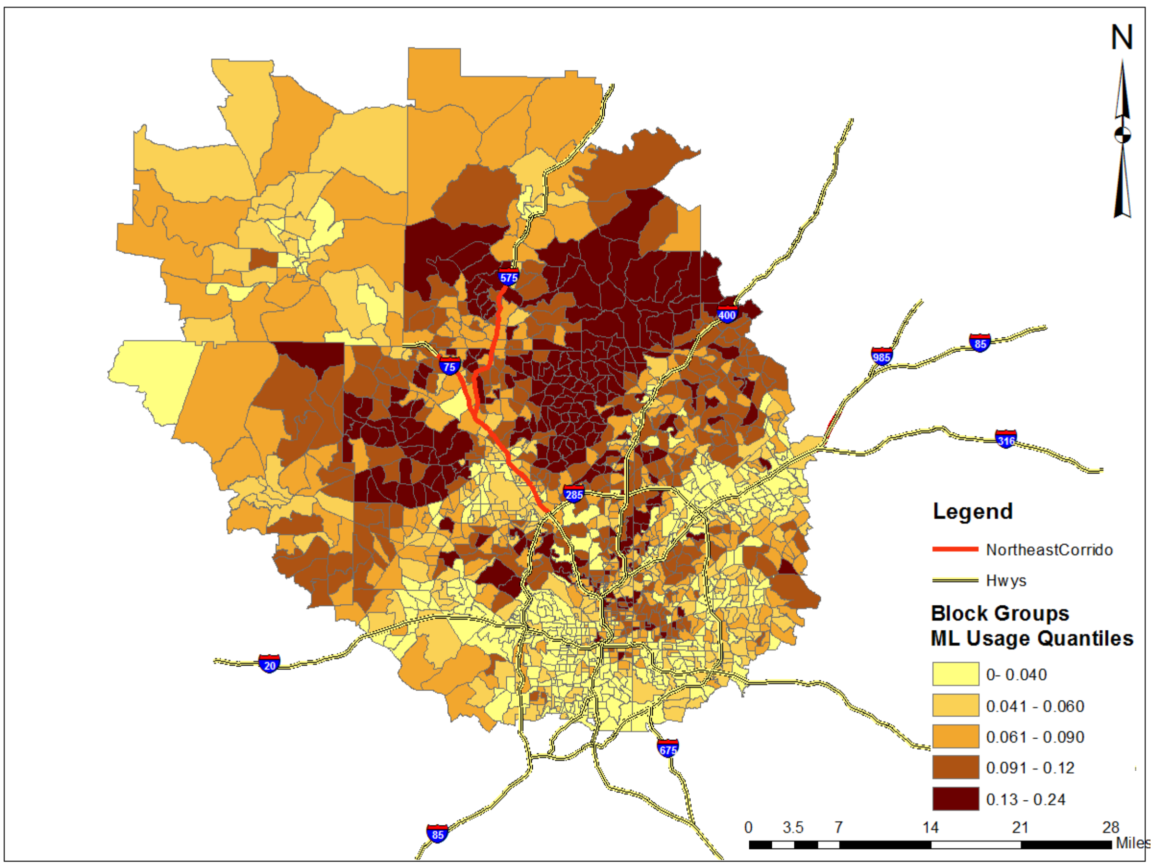


Figure 111 Northeast Corridor Block Group Level HOT Lane Usage Estimates

CHAPTER 10

CONCLUSION

Congestion pricing is a fairly recent strategy for congestion mitigation and travel time reduction. High Occupancy Toll (HOT) lanes, as a form of congestion pricing, manage capacity more efficiently and provide consumers with a travel choice that has a shorter and more reliable travel time. Investigating the impact of users' socio-spatial characteristics and their travel behavior toward HOT lane is important for policy decisions concerning future HOT lane investments and developments (tolling and revenue studies), travel demand modeling, and responding to socioeconomic concerns.

How socio-spatial characteristics impact the users travel behavior toward HOV-to-HOT conversion is the main research question of this study. This research is a case study of the conversion of High Occupancy Vehicle (HOV) lane to HOT lane, implemented in 15.5 miles of Atlanta I-85 on Oct, 1 2011.

Thus far, the travelers' response toward managed lanes is often estimated using stated preference or travel diary surveys, of small percent of the population, which are expensive, time-consuming, and labor-intensive. To minimize the cost and maximize the size of the collected data, an innovative and relatively inexpensive modeling framework for socioeconomic analysis of managed lanes has been developed and implemented. Instead of surveys, this research is based on one and a half million license plates, matched to household locations (using vehicle registration database), collected over two-year study period before and after HOV-to-HOT conversion.

Socioeconomic data supplement the household corridor usage information derived from license plate observations. Socioeconomic data are retrieved from two sources: block group level American Community Survey data, and household level marketing data. Marketing data provide very detailed household and individual level attributes with significant low amount of cost (¢10 per household), compared to travel surveys which cost about $200/household. Marketing data, in conjunction with associated trip data, has been introduced as an alternative for conducting travel behavior studies.

10.1. Research Findings

The quality of marketing data has been evaluated through comparative analysis with self-reported survey data. Income and ethnicity, dwelling type and ownership have better quality, comparing marketing data to self-reported survey data; while household size variables have lower quality (due to discrepancies in number of adults) of which number of children has better quality. Accuracy of vehicle ownership, assessed by comparing registration to self-reported data, is acceptable, considering the high level of instability of this variable within households across time. Whereas the accuracy and coverage of marketing data are not as good as survey data, large enough sample of marketing data could potentially cancel out the errors across the user groups as indicated by the fact that the distributions of the observations were quite similar.

Recent advances in trip data collection via ALPRs, tag readers and cell phones tracking and handheld surveys will likely replace travel diaries in the future and highlight the need for a supplementary socioeconomic data source. Marketing data provide very detailed household and individual level attributes at a significantly lower cost (and much larger sample sizes), which is close to impossible to be collected in surveys.

Combining the commuters' frequencies of corridor use, across the HOV, HOT and general purpose lanes, before and after the conversion, with the socioeconomic data produced the dataset for analysis. The sensitivity of license plate data collection with respect to time, day and location of data collection was assessed to help future studies with more effective deployments. The sensitivity of demographic data is lowest across days of week and higher across time of day and site of data collection within the corridor.

This study also examined the application of vehicle value, which is less expensive and more convenient to collect, as a proxy for household income. The analysis demonstrated that the average vehicle value in the HOT lane is significantly higher, about $2,100 (23%), and the average vehicle model year is about one year newer, compared to the general purpose lanes. Furthermore, of 23% difference in vehicle value between HOT and GP lanes, 13% is associated with a difference (increase) in model year, and 10% is associated with changes in vehicles make/model rankings. Moreover, HOT lanes are Accord, Civic, Camry, and F150 lanes; just like the GP lanes; hence, this study once again rejects the concept of "Lexus Lane".

Descriptive statistics and visualization techniques have been used to compare and understand the socioeconomic differences between different groups of corridor users using both block group level and household level data.

10.1.1. Descriptive Statistics

At the household level (using marketing data), two series of analysis were conducted. The first series of analysis included all the observation to compare socioeconomic attributes across the lanes at the household level, before and after the conversion. The second series of analysis included only frequent corridor users to establish user markets and compare socioeconomic attributes across the markets.

10.1.1.1. Descriptive Statistics across the Lanes

In terms of income, HOT lane user average household income is about 15% higher than users of adjacent GP lanes and HOV lane. The HOT lane represent a different commuter profile (compared to other lanes) with a shift of about 10% of users from low/medium income groups (less than $75,000) into the very high income group (more than $120,000). This was not surprising, as higher income households were expected to use the lane more frequently than lower income households (Ross et al., 2008).

In terms of vehicle ownership, HOV lane has the highest average vehicle ownership which accounts for 5% difference compared to adjacent GP lanes. HOV lane represents 15% more households with four or more vehicles and 8% fewer households with two vehicles compared to the adjacent general purpose lanes. HOT lane users represent 22% fewer households with one vehicle ownership and 9% more households with two vehicle ownership compared to the adjacent general purpose lanes.

In terms of number of adults, HOT lane represents a different distribution by 13% fewer households with one adult and 11% more households with two adults compared to the adjacent general purpose lanes. In terms of number of children, average number of children along the HOT lane is very close to one and is 27% higher compared to adjacent GP lanes and HOV lane. HOT lane represents a different distribution, with 11% fewer households with no children, about 12% more households with one or two children, and 36% more households with three or more children compared to the adjacent general purpose lanes. Hence, large families with multiple children do appear to use the HOT lanes more frequently.

Ethnicity is one of the key factors in the analysis. The original HOV lane represented 50% more Asian and 33% more Hispanic households, and accordingly 8% fewer White households compared to the adjacent general purpose lanes. On the other hand, HOT lane represents 8% more White households, and 28% fewer African-American, 33% fewer Hispanic, and 12% fewer Asian households.

In terms of education level, HOV lane represents 16% more households with high school (or lower) degree, whereas HOT lane represents 14% fewer households with the same education level compared to the adjacent general purpose lanes. Furthermore, HOT lane represents 13% more households with Bachelor or more education level. These variables are somewhat correlated with income and may be related to job classification.

In terms of marital status, HOT lane represent 14% more married households (also correlated with HH size as discussed earlier). Head of household age is not substantially changing across the lane. The only slight difference is that HOT lane has 9% more users with head of household age of 35-44 years old and accordingly 12% fewer users with head of household age of 55-64 years old. In terms of head of household gender across the lanes, no meaningful difference could be observed.

In terms of home ownership, the HOT lane has 44% fewer renters compared to the adjacent general purpose lanes. No significant change could be observed comparing HOV and HOT lanes to adjacent general purpose lanes regarding dwelling type. Lastly, no significant change can be observed across the lanes in terms of length of residence.

10.1.1.2. Descriptive Statistics across the Markets

This series set of analysis included only top 5% frequent households (representing 42% of collected license plates) and established four user markets based on their travel choice:

- *GP Market:* Commute via HOV lane 20% or less before, and commute via HOT lane 20% or less after
- *HOV Market: C*ommute via HOV lane more than 20% before, but commute via HOT lane 20% or less after
- *HOT Market: C*ommute via HOV lane 20% or less before, but commute via HOT lane more than 20% after
- *HOV/HOT* Market: Commute via HOV lane more than 20% before, and commute via HOT lane more than 20%

The HOT and HOV/HOT markets included high-income users (more than $120,000 household annual income), almost 50% more frequently than the very-low-income and low-income users. However, the GP and HOV markets contain high-income users almost 10% less than very-low-income and low-income users.

Surprisingly, 37% of HOV market consists of users from households owning four or more vehicles, compared to between 27% and 30% for the other markets. The household size profiles do not illustrate as much difference as other attributes.

Regarding ethnicity, the proportion of Hispanic and Asian population in the HOV market is about two times (or more) larger than GP and HOV/HOT markets and about three times larger than HOT market. Furthermore, proportion of the Asian population in the HOV market is about two times larger than in the HOT and GP markets.

Although household gender doesn't show significant and meaningful difference across the markets, head of household age varied significant across the markets. HOT and HOV/HOT markets include 12% more users with head of household in the age range of 35-44, compared to GP and HOV markets. Regarding education, the differences across the markets are also significant. Compared to the GP and HOV markets, 21% increase in users with associate/graduate degrees in HOV/HOT market and 14% increase in HOT market were observed.

Despite the observed variations across the markets and lanes demographic groups, all the demographic groups are well representing across all the markets. This implies that the HOV-to-HOT conversion did not remove any demographic group from use of the managed lane, although it certainly had impacts on frequency of managed lane and corridor use.

10.1.2. Modeling

In terms of modeling, this study developed six models at two analytical levels: primary aggregated level (block group level) and advanced disaggregated level (household level). The advantages of the block group level models are lower cost, and publically available socioeconomic data, and the disadvantage is lower predictive power. The advantage of household level models is significantly higher predictive power, considering the higher cost of data collection and marketing data acquisition. Household level models accuracy increases as the sample size and resulting cost of data increase.

10.1.2.1. Block Group Level Modeling

Using the block group level HOT lane model, income, vehicle ownership, gender (percent of female residents), and percent of home workers have positive impacts, while household size, age, education and percent of drivers (for commute) have negative impacts on block group HOT lane usage. HOV lane usage has been modeled with spatial lag model, because of the high spatial correlation between the block groups with respect to the HOV usage. Controlling for HOV lane usage spatial autocorrelation, income, education, percent of drivers (for commute), percent of home workers, and commute travel time have negative impacts while household size and age have positive impact.

It should be noted that HOT usage model has substantially better goodness of fit compared to the HOV usage model. In other words, socioeconomic variables are more effective in modeling HOT lane use compared to HOV lane use (carpool formation). Detailed survey data from carpoolers could potentially improve our knowledge about the underlying reasons regarding HOV lane usage, in the future.

10.1.2.2. Household Level Modeling

Using frequent corridor commuters and household level marketing data, two logistic regression models were developed to investigate the impact of socioeconomic attributes on users' assignment across the markets.

The first model predicts HOV lane frequent users' choice to either shift out of the HOV lane into the GP lanes (HOV market) or to continue using the HOT lane after conversion (HOV/ HOT market). Households with Hispanic, Asian, and African-American ethnicity increase the chance of switching to GP lanes, whereas high income, high vehicle ownership, and children (one and three) decrease the chance of switching to GP lanes and increase the chance of HOV/HOT market.

The second model predicts the GP lane users' choice to either continue using the GP lane (GP market) or switch from the GP lane into the HOT lane after conversion (HOT market). Households with Hispanic, African-American and Asian ethnicity, college degrees, high length of residence, and older head of household decrease the chance of shifting to the HOT market, whereas households with high income, home ownership, female head of household, and Bachelor (or higher) degree increase the chance of switching to the HOT market.

The last developed models (count models) predict number of HOV lane and HOT lane use per household as a function of socioeconomic attributes and general corridor use frequency.

Households with Hispanic and Asian ethnicity, older head of households, two (or more) adults and no children, high vehicle ownership, low education level, male head of household, and who are owners of single family units are making relatively more HOV trips.

Households with home ownership, higher income, White ethnicity, younger head of household, one adults/two married adults and one or more children, two (or more) vehicle ownership, and higher education levels are making relatively more HOT trips.

Count models are perhaps the most valuable models developed in this study; because they use all of the observed license plate data and have the highest goodness of fit (0.53 for HOV model and 0.68 for HOT model). Count models have potentially very high applicability for use in traffic and revenue studies, where predicted traffic counts by socioeconomic market can be combined with demand elasticity research results to predict demand and toll and revenues.

Generally, the impacts of income, home ownership and ethnicity (Hispanic/Asian/African-American) are the highest in these models. The fact that income and home ownership are significant is obvious and intuitive, considering the pricing scheme of the conversion. However, the impact of ethnicity after controlling for income is very interesting and has not been identified in any previous studies. One potential reason might be the fact that these specific ethnic groups might be more hesitant to acquire transponders.

Lastly, the presence of correlations between socioeconomic attributes, present or absent from the model, could result in an overestimate or underestimate of variable impacts. This is potentially true with respect to ethnicity. For example, carpooling rates are relatively different across ethnic groups. However, importing the interactions between variables to the model could help better understand the impact of socioeconomic variables and is in scope of work for further expansion of the models in future.

10.1.3. Spatial Analysis

GIS raster analysis methods have been utilized to visualize and quantify the impact of the HOV-to-HOT conversion on corridor commutershed. The HOT lane commutershed is smaller than the HOV lane commutershed and the general purpose lane commutershed has expanded after the conversion (perhaps in part resulting from longer distance commuters switching from the HOV lane to general purpose lane, as well as the addition of new long-distance commuters). However, the amount of commutershed expansion by general purpose lanes dominates the amount of retraction produced by HOT lane, causing an overall expansion in the corridor commutershed.

In summary, the corridor usage slightly reduced in substantial portion of the corridor commutershed during the AM peak period. The reduction in general purpose lanes usage is mainly concentrated around the close northern part of the corridor (mainly around Lawrenceville area), while the reduction in HOT lane usage is mainly along the corridor, spatially correlated with low income and Hispanic population residential areas. Additionally, the substantial positive impact has been observed in HOT lane usage in far northern part of the corridor, spatially correlated with high income population areas.

The PM peak commutershed is generally more expanded toward the CBD areas, considering all the non-commuting (shopping, leisure, and etc.) trips in the afternoon. In addition to the quite similar reduction in corridor usage to AM peak, the majority of the positive change in corridor usage has been observed in the afternoon peak all over the commutershed, causing the overall corridor commutershed to expand.

10.2. Limitations

As discussed earlier, there are other elements beside socioeconomic characteristics that impact users' travel choice. Time of day, trip purpose, trip length, vehicle occupancy, traffic condition, personal attitudes, toll amount, and willingness to pay are probably the major determining elements in users' travel choice. Considering the forecasting purpose of this study, only the socioeconomic attributes have been considered here. Therefore, the models developed in this study can only be applicable for macro forecasting and planning purposes and they are not appropriate for choice modeling at trip level.

The assumption that the registration location of the license plate is same as the household that the vehicle user is can be questioned. This concern can be addressed partially by excluding all the households out of the 95% or 99% commutershed; however, this concern may also exist inside the commutershed to some extent.

Although using a third party data is convenient and relatively inexpensive, the marketing data and registration data applied in this study are not 100% accurate and up-to-date. Furthermore, there is always a time lag between observing the license plate in the corridor, matching to registration database and purchasing household demographic data. These time lags create some concerns about the accuracy of the data.

Lastly, all the analyses and models illustrated in this study are based on a special form of congestion pricing implementation (HOV2+ to HOT3+ conversion) without the addition of extra capacity to the corridor. Moreover, the I-85 corridor commutershed consist of various ethnic groups. Future application of the study results should consider the potential differences between the planned project specifications and the I-85 corridor and consequently, the potential impacts on the results.

10.3. Future Work

Developing a calculator tool which can automate the process of applying he developed models in this study for future tolling and revenue studies is the next step of research in this area. This calculator should be able to assess the amount of required data to achieve a certain standard error of the predictions.

Furthermore, the addition of trip level characteristics such as speed, toll amount, and vehicle occupancy can increase the prediction power of the models and conduct forecasting at both macro and micro levels. None of the models developed in this study include toll price. While the sensitivity to toll pricing is inherent in the models, the price relative to disposable income may be a very important factor in the next generation of the models. The future models refine the developed household-level models to trip-level, and predict households travel behavior as a function of toll price, level of congestion, time of day, and any available trip level attribute in addition to socioeconomic attributes.

A follow-up travel survey could potentially clarify the underlying reasons for some counter-intuitive findings such as high vehicle ownership for HOV market, low impact of household size variables, very low use of HOT lane by Hispanic and Asian population, and decrease in corridor use around Lawrenceville area.

While the sample for this dataset is very large and the study period covers one year before and one year after the HOV to HOT lane conversion, still there are some concerns about the quality of the data. The most comprehensive data source for this type of analysis is instrumented fleet data which incorporate all the users and trip attributes. Because we didn't have the opportunity to collect this type of data, license plate data have been collected which does not cover all the aspects of transportation decision making process. Meanwhile, application of advanced technologies such as Automatic License Plate Readers, Cell Phone Location Data and/or Travel Apps, RFID tag reads could enhance the data quality and decrease the cost for similar studies in the future.

APPENDIX A

MARKETING DATA VARIABLES

Credit bureau data covers all detail aspects of the household information. Following is a summarized list of information available for purchase provided by typical marketing agencies.

- Geographic
  - Address
  - Latitude and Longitude

- Demographic
  - Household
    - Household Size
    - Family Composition
    - House Ownership Status
    - Dwelling Type
    - Length of Residence
    - Lifestyle (NICHES)
  - Residence
    - Age
    - Gender
    - Income
    - Ethnicity
    - Relationship
    - Occupation
    - Education
    - Marital Status

- Financial
  - Household Income
  - House
    - Living Area Square Feet
    - Property Lot Size in Acre
    - Home Market Value
    - Year Home Built
  - Automobiles
    - Number of Cars ,Trucks, MCs and RVs

- Market Value
- Number of Registered Vehicles
- Number of Currently Owned Vehicles
- Number of Currently Leased Vehicles

Table 36 I-85 Corridor Users Ethnicity Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
White	159,278	66.0	68.3	68.3
African American	24,754	10.3	10.6	78.9
Hispanic	17,970	7.4	7.7	86.6
Asian	15,585	6.5	6.7	93.3
Other	15,636	6.5	6.7	100.0
Total	233,223	96.6	100.0	
Missing	8,243	3.4		
Total	241,466	100.0		

Table 37 I-85 Corridor Users Gender Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
Male	121,466	50.3	54.3	54.3
Female	102,206	42.3	45.7	100.0
Total	223,672	92.6	100.0	
Missing	17,794	7.4		
Total	241,466	100.0		

Table 38 I-85 Corridor Users Head of HH Age Frequency Distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
1	**18-24 years old**	2,251	.9	.9	.9
2	**25-34 years old**	35,284	14.6	14.6	15.6
3	**35-44 years old**	74,469	30.8	30.9	46.4
4	**45-54 years old**	60,168	24.9	24.9	71.4
5	**55-64 years old**	41,086	17.0	17.0	88.4
6	**65-74 years old**	19,595	8.1	8.1	96.6
7	**75+ years old**	8,302	3.4	3.4	100.0
	Total	241,155	99.9	100.0	
Missing	311	.1			
Total	241,466	100.0			

Table 39 I-85 Corridor Users Head of HH Age (Short Format) Frequency Distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
1	**18-34 years old**	2,251	.9	.9	.9
2	**35-54 years old**	35,284	14.6	14.6	15.6
3	**55+ years old**	74,469	30.8	30.9	46.4
Total		241,466	100.0		

Table 40 I-85 Corridor Users Marital Status Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
Married	114,685	47.5	48.6	48.6
Single	121,238	50.2	51.4	100.0
Total	235,923	97.7	100.0	
Missing	5,543	2.3		
Total	241,466	100.0		

Table 41 I-85 Corridor Users Presence of Children Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
No	140,449	58.2	58.2	58.2
Yes	100,706	41.7	41.8	100.0
Total	241,155	99.9	100.0	
Missing	311	.1		
Total	241,466	100.0		

Table 42 I-85 Corridor Users Length of Residence Frequency Distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
1	**0-6 Monthes**	10,521	4.4	4.4	4.4
2	**7-12 Monthes**	13,578	5.6	5.6	10.0
3	**1-2 Years**	16,012	6.6	6.6	16.6
4	**3-5 Years**	46,321	19.2	19.2	35.8
5	**6-10 Years**	69,451	28.8	28.8	64.6
6	**11-15 Years**	44,039	18.2	18.3	82.9
7	**16-20 Years**	19,443	8.1	8.1	91.0
8	**20+ Years**	21,790	9.0	9.0	100.0
Total		241,155	99.9	100.0	
Missing		311	.1		
Total		241,466	100.0		

Table 43 I-85 Corridor Users Living Area (Square Feet) Frequency Distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
1	**1-749**	345	.1	.3	.3
2	**750-999**	1,440	.6	1.2	1.5
3	**1000-1249**	6,965	2.9	5.9	7.4
4	**1250-1499**	11,973	5.0	10.1	17.4
5	**1500-1749**	14,560	6.0	12.2	29.6
6	**1750-1999**	13,010	5.4	10.9	40.6
7	**2000-2499**	24,228	10.0	20.4	60.9
8	**2500-2999**	21,154	8.8	17.8	78.7
9	**3000-3499**	12,714	5.3	10.7	89.4
10	**3500-3999**	6,211	2.6	5.2	94.6
11	**4000-4999**	4,297	1.8	3.6	98.2
12	**5000-5999**	1,297	.5	1.1	99.3
13	**6000-6999**	486	.2	.4	99.7
14	**7000 and more**	358	.1	.3	100.0
Total		119038	49.3	100.0	
Missing		122428	50.7		
Total		241466	100.0		

Table 44 I-85 Corridor Users Dwelling Type Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
Single Family Unit	213,811	88.5	89.9	89.9
Multi Family Unit / Condo	23,965	9.9	10.1	100.0
Total	237,776	98.5	100.0	
Missing	3,690	1.5		
Total	241,466	100.0		

Table 45 I-85 Corridor Users Home Ownership Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
Renter	26,316	10.9	11.0	11.0
Owner	211,895	87.8	89.0	100.0
Total	238,211	98.7	100.0	
Missing	3,255	1.3		
Total	241,466	100.0		

Table 46 I-85 Corridor Users Education Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
High School or lower	66,485	27.5	27.6	27.6
College	84,143	34.8	34.9	62.5
Bachelor or higher	90,404	37.4	37.5	100.0
Total	241,032	99.8	100.0	
Missing	434	.2		
Total	241,466	100.0		

Table 47 I-85 Corridor Users Household Vehicle Ownership Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
1	49,973	20.7	20.7	20.7
2	76,030	31.5	31.5	52.2
3	50,939	21.1	21.1	73.3
4	31,854	13.2	13.2	86.5
5	17,819	7.4	7.4	93.8
6	9,833	4.1	4.1	97.9
7	5,018	2.1	2.1	100.0
Total	241,466	100.0	100.0	

Table 48 I-85 Corridor Users Household Size Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
1	88,457	36.6	36.7	36.7
2	51,664	21.4	21.4	58.1
3	46,765	19.4	19.4	77.5
4	25,626	10.6	10.6	88.1
5	13,930	5.8	5.8	93.9
6	7,171	3.0	3.0	96.9
7	3,674	1.5	1.5	98.4
8	1,916	.8	.8	99.2
9	1,952	.8	.8	100.0
Total	241,155	99.9	100.0	
Missing	311	.1		
Total	241,466	100.0		

Table 49 I-85 Corridor Users Household Number of Children Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
0	140,449	58.2	67.7	67.7
1	31,698	13.1	15.3	83.0
2	15,850	6.6	7.6	90.7
3	9,140	3.8	4.4	95.1
4	5,015	2.1	2.4	97.5
5	2,587	1.1	1.2	98.7
6	1,390	.6	.7	99.4
7	773	.3	.4	99.8
8	331	.1	.2	99.9
9	111	.0	.1	100.0
Total	207,344	85.9	100.0	
Missing	34,122	14.1		
Total	241,466	100.0		

Table 50 I-85 Corridor Users Household Number of Adults Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
1	95525	39.6	39.6	39.6
2	100072	41.4	41.5	81.1
3	32536	13.5	13.5	94.6
4	10243	4.2	4.2	98.8
5	2779	1.2	1.2	100.0
Total	241155	99.9	100.0	
Missing	311	.1		
Total	241466	100.0		

Table 51 I-85 Corridor Users Household Annual Income Frequency Distribution

	Frequency	Percent	Valid Percent	Cumulative Percent
Less than $15,000	7,957	3.3	3.3	3.4
$ 15,000 - $ 19,999	6,634	2.7	2.7	6.2
$ 20,000 - $ 29,999	24,750	10.2	10.2	16.4
$ 30,000 - $ 39,999	34,122	14.1	14.1	30.6
$ 40,000 - $ 49,999	33,851	14.0	14.0	44.6
$ 50,000 - $ 74,999	58,917	24.4	24.4	69.0
$ 75,000 - $ 99,999	33,847	14.0	14.0	83.0
$ 100,000 - $124,999	17,759	7.4	7.4	90.3
$ 125,000 - $149,999	14,646	6.1	6.1	96.4
$ 150,000 - $174,999	4,236	1.8	1.8	98.2
$ 175,000 - $199,999	1,779	.7	.7	98.9
$ 200,000 - $249,999	1,419	.6	.6	99.5
$ 250,000 or more	1,238	.5	.5	100.0
Missing	311	.1	.1	.1
Total	241,466	100.0	100.0	

APPENDIX B

LOGISTIC MODELS OUTPUTS

Table 52 HOV Users Response Logit 1 Model: HOV Market vs. HOV/HOT Market (base choice)

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test	df	Sig.	Exp (B)	95% Wald Confidence Interval for Exp (B)	
			Low	Up	Wald Chi-Sq				Low	Up
Intercept	1.37	0.57	0.25	2.49	5.75	1.00	0.02	3.94	1.28	12.09
Income_Ln	-0.22	0.12	-0.44	0.01	3.41	1.00	0.07	0.81	0.64	1.01
Vehicle=1	0	.	.	.	.	.	.	1.0	.	.
Vehicle=2	-0.51	0.18	-0.88	-0.15	7.73	1.00	0.01	0.60	0.42	0.86
Vehicle=3	-0.47	0.19	-0.85	-0.09	5.82	1.00	0.02	0.63	0.43	0.92
Vehicle=4+	-0.19	0.19	-0.56	0.17	1.08	1.00	0.30	0.82	0.57	1.19
Adults=1	0	.	.	.	.	.	.	1.00	.	.
Adults=2	0.13	0.18	-0.22	0.47	0.51	1.00	0.47	1.14	0.80	1.61
Adults=3	0.36	0.21	-0.04	0.76	3.06	1.00	0.08	1.43	0.96	2.14
Children=1	0	.	.	.	.	.	.	1.00	.	.
Children=2	-0.20	0.17	-0.52	0.12	1.45	1.00	0.23	0.82	0.59	1.13
Children=3	-0.02	0.20	-0.42	0.38	0.01	1.00	0.92	0.98	0.66	1.46
Children=4+	-0.51	0.20	-0.91	-0.11	6.32	1.00	0.01	0.60	0.40	0.89
Ethnicity=White	0	.	.	.	.	.	.	1.00	.	.
Ethnicity=African	0.50	0.29	-0.07	1.08	2.94	1.00	0.09	1.66	0.93	2.95
Ethnicity=Hispanic	0.83	0.18	0.48	1.18	21.92	1.00	0.00	2.29	1.62	3.25
Ethnicity=Asian	0.29	0.17	-0.03	0.62	3.10	1.00	0.08	1.34	0.97	1.85
Ethnicity=Other	0.00	0.21	-0.41	0.41	0.00	1.00	1.00	1.00	0.66	1.51
Education=HS-	0	.	.	.	.	.	.	1.00	.	.
Education=College	0.17	0.16	-0.14	0.49	1.13	1.00	0.29	1.19	0.87	1.63
Education=BS+	-0.02	0.17	-0.35	0.31	0.02	1.00	0.89	0.98	0.70	1.36
Marital Stat=Married	0	.	.	.	.	.	.	1.00	.	.
Marital Status=Single	0.10	0.19	-0.27	0.46	0.27	1.00	0.60	1.10	0.76	1.59
Age = 18-34 yrs old	0	.	.	.	.	.	.	1.00	.	.
Age = 35-54 yrs old	-0.14	0.17	-0.48	0.20	0.69	1.00	0.41	0.87	0.62	1.22
Age=55+ yrs old	-0.24	0.21	-0.65	0.17	1.30	1.00	0.25	0.79	0.52	1.19
Gender= Male	0	.	.	.	.	.	.	1.00	.	.
Gender= Female	0.01	0.11	-0.20	0.23	0.01	1.00	0.91	1.01	0.82	1.26
Renter	0	.	.	.	.	.	.	1.00	.	.
Owner	-0.13	0.29	-0.71	0.44	0.21	1.00	0.65	0.87	0.49	1.55
Dwelling Type=SFDU	0	.	.	.	.	.	.	1.00	.	.
Dwelling Type=MFDU	-0.08	0.26	-0.59	0.43	0.10	1.00	0.75	0.92	0.55	1.54
Length of Residence	0.02	0.04	-0.06	0.10	0.22	1.00	0.64	1.02	0.94	1.11

Table 53 HOV Users Response Logit 2 Model: HOV Market vs. HOV/HOT Market (base choice)

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test	df	Sig.	Exp (B)	95% Wald Confidence Interval for Exp (B)	
			Low	Up	Wald Chi-Sq				Low	Up
Intercept	0.84	0.50	-0.15	1.83	2.79	1.00	0.10	2.32	0.87	6.22
Income_Ln	-0.23	0.10	-0.43	-0.04	5.4	1.00	0.02	0.79	0.65	0.96
Vehicle=1	0	.	.	.	.	.	.	1.00	.	.
Vehicle=2	-.51	.181	-.87	-.16	8.11	1	.00	.59	.41	.85
Vehicle=3	-.45	.190	-.82	-.08	5.73	1	.01	.63	.43	.92
Vehicle=4+	-.17	.182	-.53	.18	0.87	1	.34	.84	.58	1.20
Children=1	0	.	.	.	.	.	.	1.00	.	.
Children=2	-.14	.146	-.43	.14	.96	1	.32	.86	.65	1.15
Children=3	.05	.177	-.29	.40	.09	1	.75	1.05	.74	1.49
Children=4+	-.45	.170	-.78	-.11	6.99	1	.00	.63	.45	.89
Ethnicity=White	0	.	.	.	.	.	.	1.00	.	.
Ethnicity=African	.53	.291	-.03	1.10	3.37	1	.06	1.70	.96	3.02
Ethnicity=Hispanic	.86	.175	.52	1.20	24.29	1	.00	2.37	1.68	3.34
Ethnicity=Asian	.30	.157	-.00	.60	3.62	1	.05	1.35	.99	1.83
Ethnicity=Other	-.00	.206	-.40	.40	.000	1	.98	.99	.66	1.49

Table 54 GP Users Response Logit 1 Model: GP Market (base choice) vs. HOT Market

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test	df	Sig.	Exp (B)	95% Wald Confidence Interval for Exp (B)	
			Low	Up	Wald Chi-Sq				Low	Up
Intercept	-3.22	0.32	-3.84	-2.60	104.58	1.00	0.00	0.04	0.02	0.07
Income_Ln	0.48	0.06	0.35	0.60	57.13	1.00	0.00	1.61	1.42	1.83
Vehicle=1	0	.	.	.	.	.	.	1.00	.	.
Vehicle=2	0.16	0.09	-0.01	0.33	3.37	1.00	0.07	1.17	0.99	1.39
Vehicle=3	0.14	0.09	-0.04	0.32	2.36	1.00	0.12	1.15	0.96	1.38
Vehicle=4+	0.12	0.09	-0.06	0.30	1.74	1.00	0.19	1.13	0.94	1.36
Adults=1	0	.	.	.	.	.	.	1.00	.	.
Adults=2	0.02	0.10	-0.17	0.21	0.06	1.00	0.81	1.02	0.85	1.24
Adults=3	-0.06	0.12	-0.28	0.17	0.25	1.00	0.62	0.94	0.75	1.18
Children=1	0	.	.	.	.	.	.	1.00	.	.
Children=2	0.09	0.08	-0.08	0.25	1.05	1.00	0.31	1.09	0.92	1.28
Children=3	-0.02	0.10	-0.23	0.18	0.05	1.00	0.82	0.98	0.80	1.20
Children=4+	0.13	0.10	-0.07	0.32	1.59	1.00	0.21	1.13	0.93	1.38
Ethnicity=White	0	.	.	.	.	.	.	1.00	.	.
Ethnicity=African	-0.15	0.14	-0.43	0.13	1.10	1.00	0.29	0.86	0.65	1.14
Ethnicity=Hispanic	-0.47	0.12	-0.70	-0.24	16.13	1.00	0.00	0.63	0.50	0.79
Ethnicity=Asian	-0.26	0.11	-0.47	-0.05	5.96	1.00	0.01	0.77	0.62	0.95
Ethnicity=Other	0.03	0.10	-0.18	0.24	0.08	1.00	0.77	1.03	0.84	1.27
Education=HS-	0	.	.	.	.	.	.	1.00	.	.
Education=College	-0.07	0.08	-0.23	0.09	0.68	1.00	0.41	0.94	0.80	1.10
Education=BS+	0.10	0.08	-0.06	0.26	1.40	1.00	0.24	1.10	0.94	1.30
Marital Status=Married	0	.	.	.	.	.	.	1.00	.	.
Marital Status=Single	0.00	0.10	-0.20	0.20	0.00	1.00	0.99	1.00	0.82	1.22
Age = 18-34 yrs old	0	.	.	.	.	.	.	1.00	.	.
Age = 35-54 yrs old	0.02	0.09	-0.15	0.19	0.04	1.00	0.85	1.02	0.86	1.21
Age=55+ yrs old	-0.11	0.11	-0.32	0.10	1.14	1.00	0.28	0.89	0.72	1.10
Gender= Male	0	.	.	.	.	.	.	1.00	.	.
Gender= Female	0.10	0.05	0.00	0.21	3.77	1.00	0.05	1.11	1.00	1.23
Renter	0	.	.	.	.	.	.	1.00	.	.
Owner	0.33	0.17	0.00	0.66	3.79	1.00	0.05	1.39	1.00	1.93
Dwelling=SFDU	0	.	.	.	.	.	.	1.00	.	.
Dwelling=MFDU	-0.12	0.13	-0.37	0.13	0.86	1.00	0.35	0.89	0.69	1.14
Length of Residence	-0.10	0.02	-0.14	-0.06	24.05	1.00	0.00	0.90	0.87	0.94

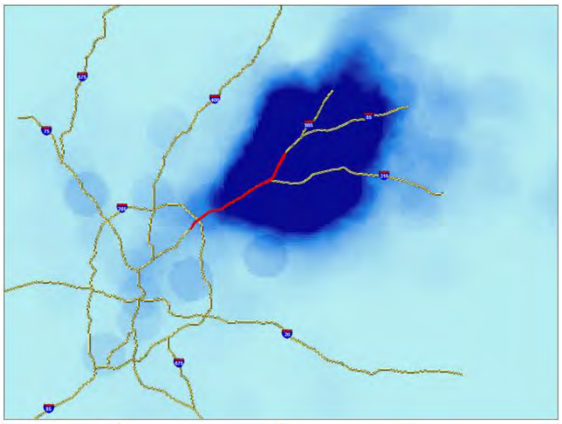
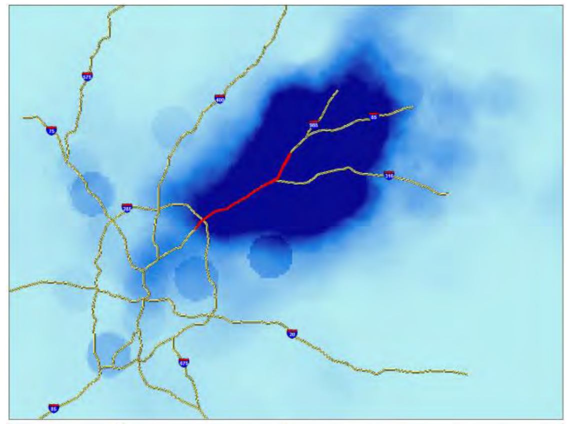
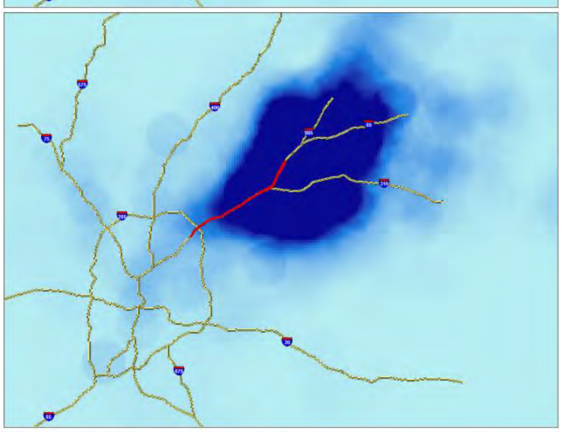
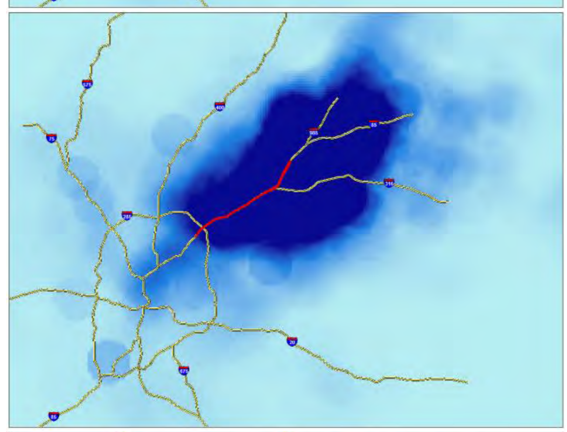
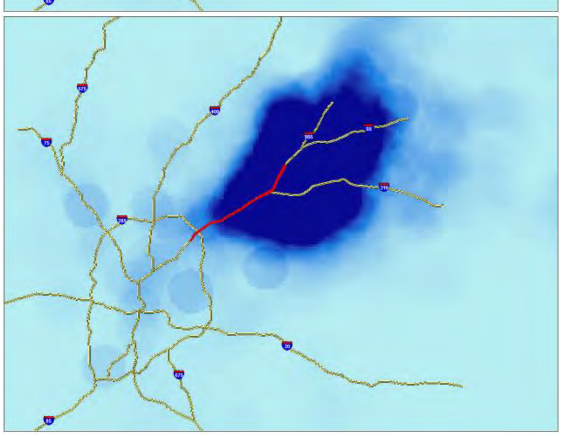
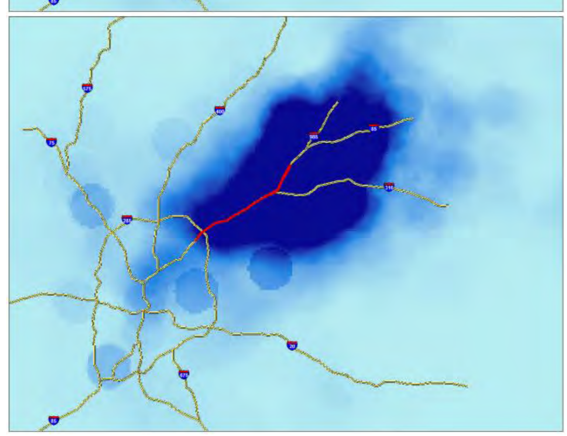
Table 55 GP Users Response Logit 2 Model: GP Market (base choice) vs. HOT Market

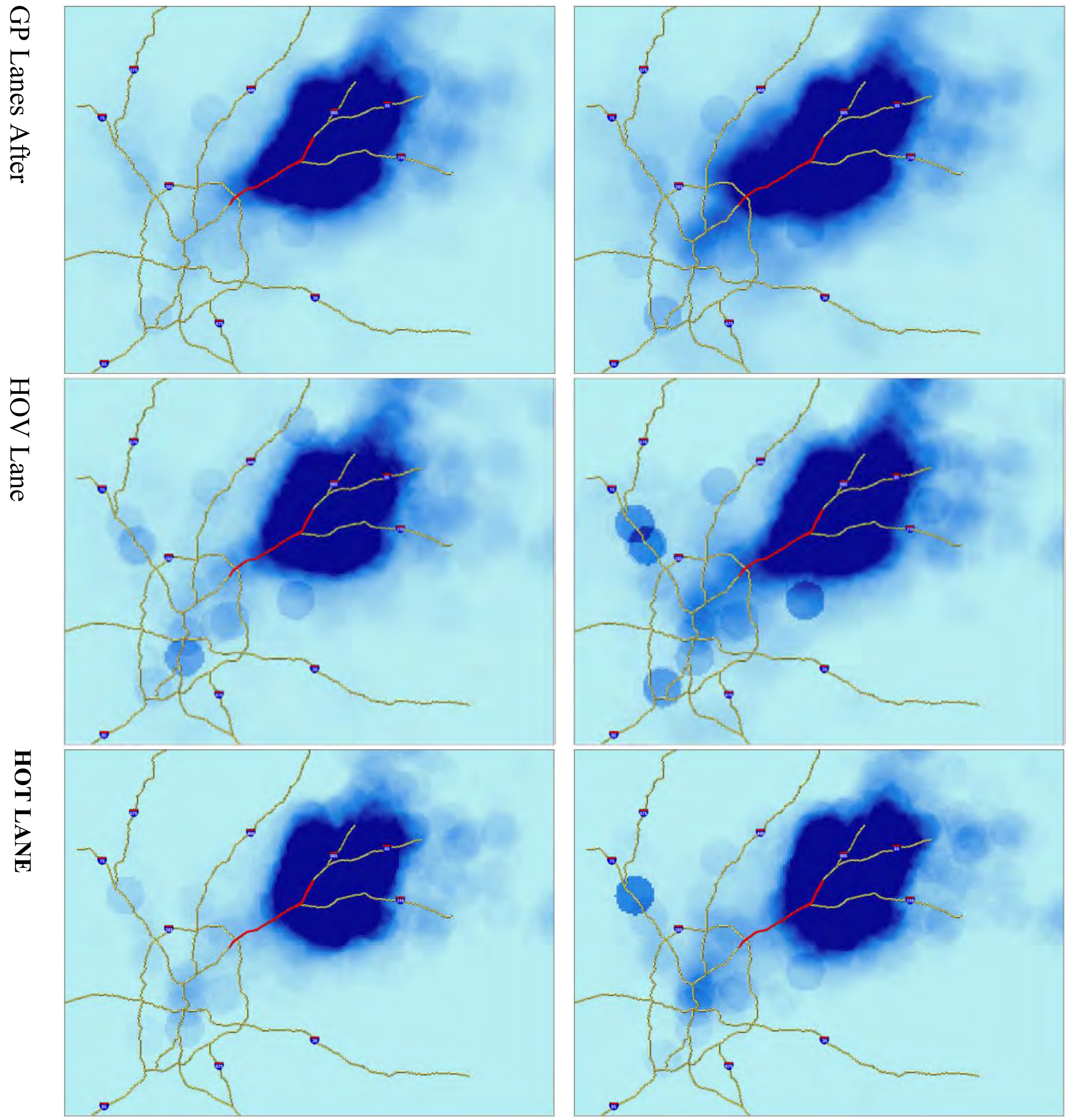
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test	df	Sig.	Exp (B)	95% Wald Confidence Interval for Exp (B)	
			Low	Up	Wald Chi-Sq				Low	Up
Intercept	-3.32	0.25	-3.81	-2.82	173.10	1.00	0.00	0.04	0.02	0.06
Income_Ln	0.51	0.06	0.40	0.63	76.86	1.00	0.00	1.67	1.49	1.87
Ethnicity=White	0	.	.	.	.	.	.	1.00	.	.
Ethnicity=African	-0.16	0.14	-0.44	0.11	1.34	1.00	0.25	0.85	0.64	1.12
Ethnicity=Hispanic	-0.47	0.12	-0.69	-0.24	16.20	1.00	0.00	0.63	0.50	0.79
Ethnicity=Asian	-0.27	0.11	-0.48	-0.07	6.65	1.00	0.01	0.76	0.62	0.94
Ethnicity=Other	0.03	0.10	-0.17	0.24	0.09	1.00	0.77	1.03	0.84	1.27
Education=HS-	0	.	.	.	.	.	.	1.00	.	.
Education=College	-0.08	0.08	-0.24	0.08	0.97	1.00	0.33	0.92	0.79	1.08
Education=BS+	0.08	0.08	-0.08	0.24	0.98	1.00	0.32	1.08	0.92	1.27
Age = 18-34 yrs old	0	.	.	.	.	.	.	1.00	.	.
Age = 35-54 yrs old	0.03	0.09	-0.14	0.20	0.11	1.00	0.74	1.03	0.87	1.22
Age=55+ yrs old	-0.15	0.10	-0.34	0.05	2.26	1.00	0.13	0.86	0.71	1.05
Gender= Male	0	.	.	.	.	.	.	1.00	.	.
Gender= Female	0.10	0.05	-0.01	0.20	3.27	1.00	0.07	1.10	0.99	1.22
Renter	0	.	.	.	.	.	.	1.00	.	.
Owner	0.42	0.16	0.12	0.73	7.41	1.00	0.01	1.53	1.13	2.08
Length of Residence	-0.10	0.02	-0.13	-0.06	24.78	1.00	0.00	0.91	0.88	0.94

APPENDIX C

COMMUTERSHED MAPS (BY TIME OF DAY)

Table 56 Fuzzy Membership Commutershed Maps (by time of day)

	AM	PM
All Lanes Before		
Il Lanes After		
GP Lanes Before		


GP Lanes After
HOV Lane
HOT LANE

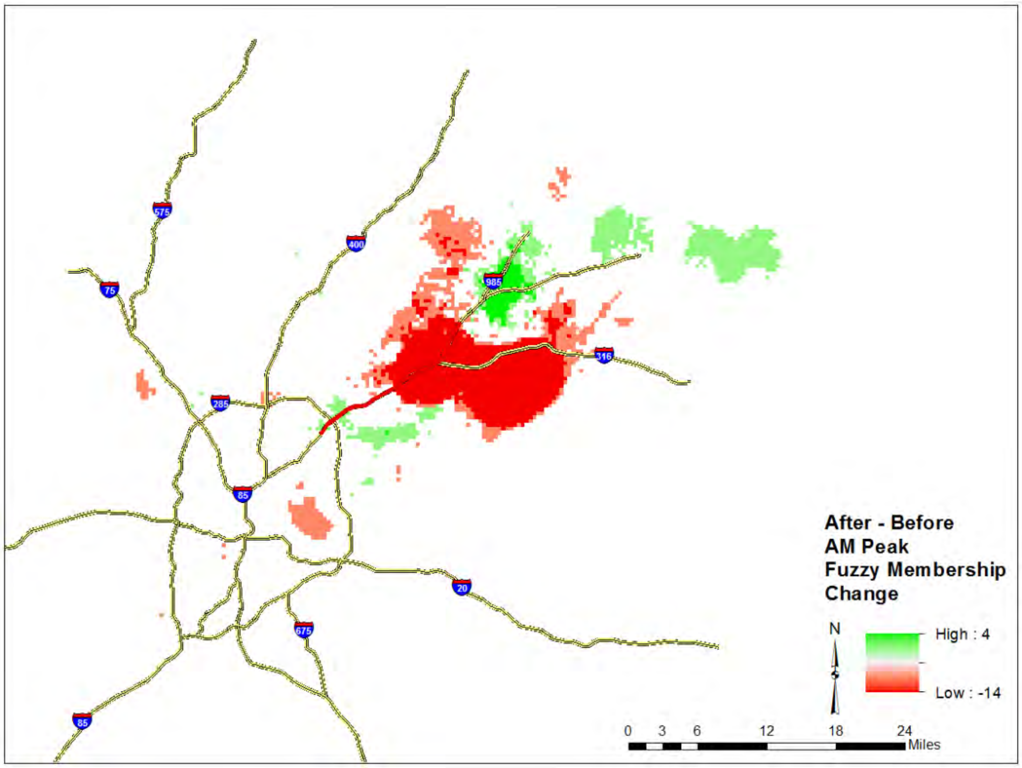


Figure 112 Before and After Conversion All Lanes Commutershed Change (AM Peak)

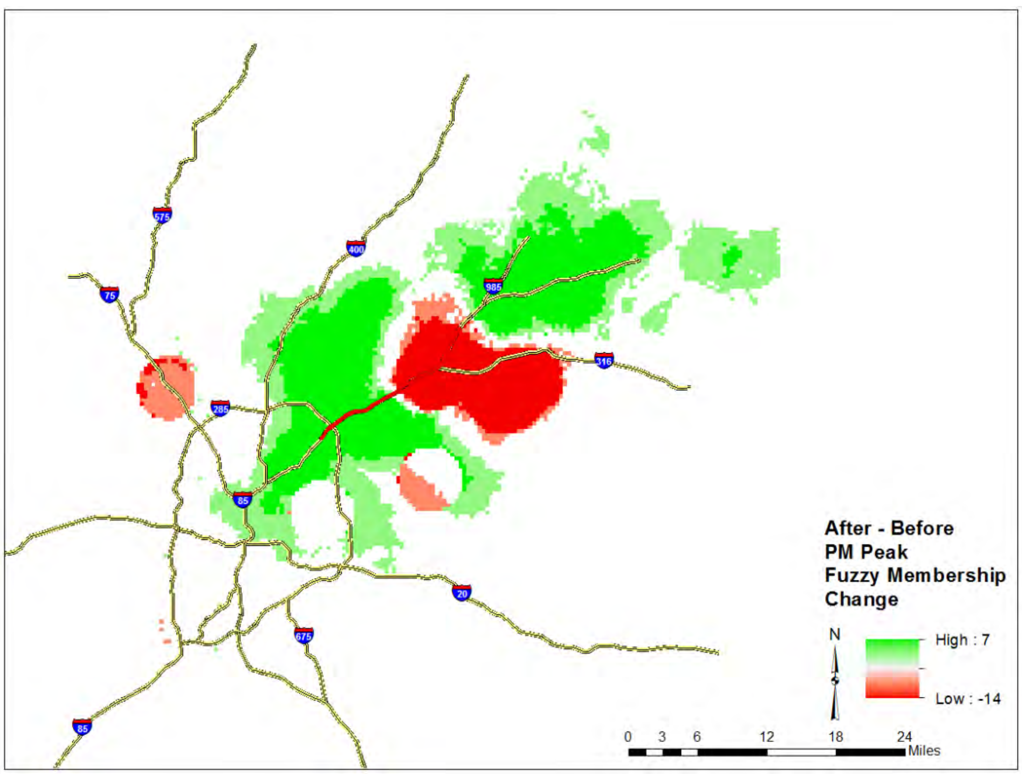


Figure 113 Before and After Conversion All Lanes Commutershed Change (PM Peak)

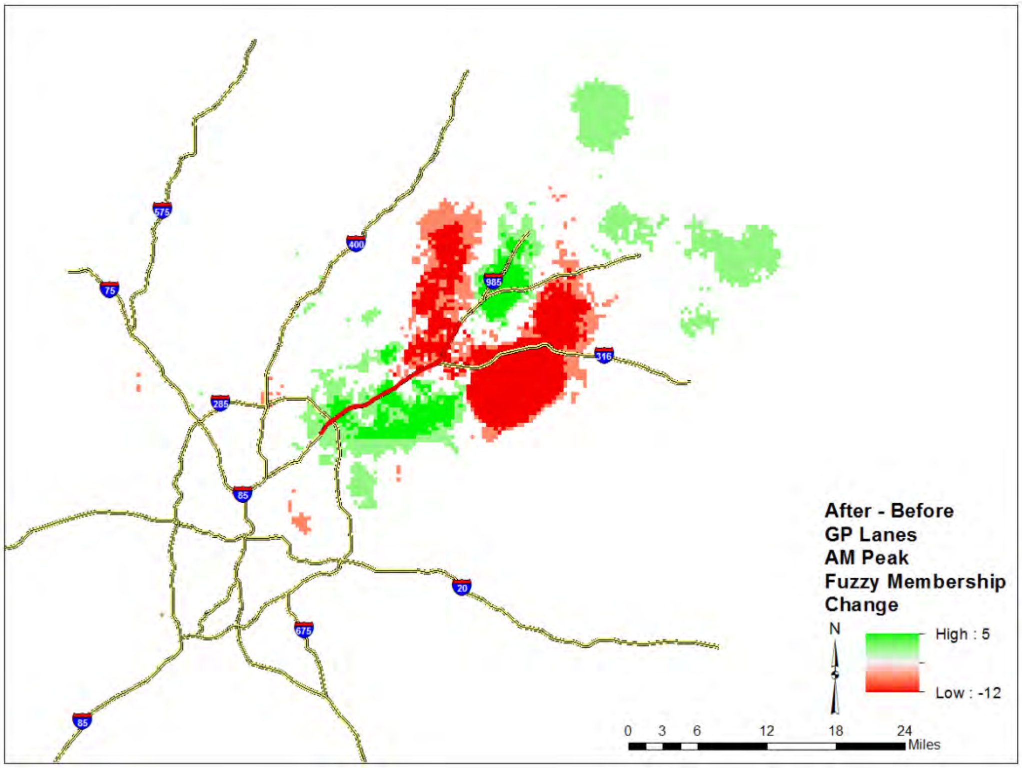


Figure 114 Before and After Conversion GP Lanes Commutershed Change (AM Peak)

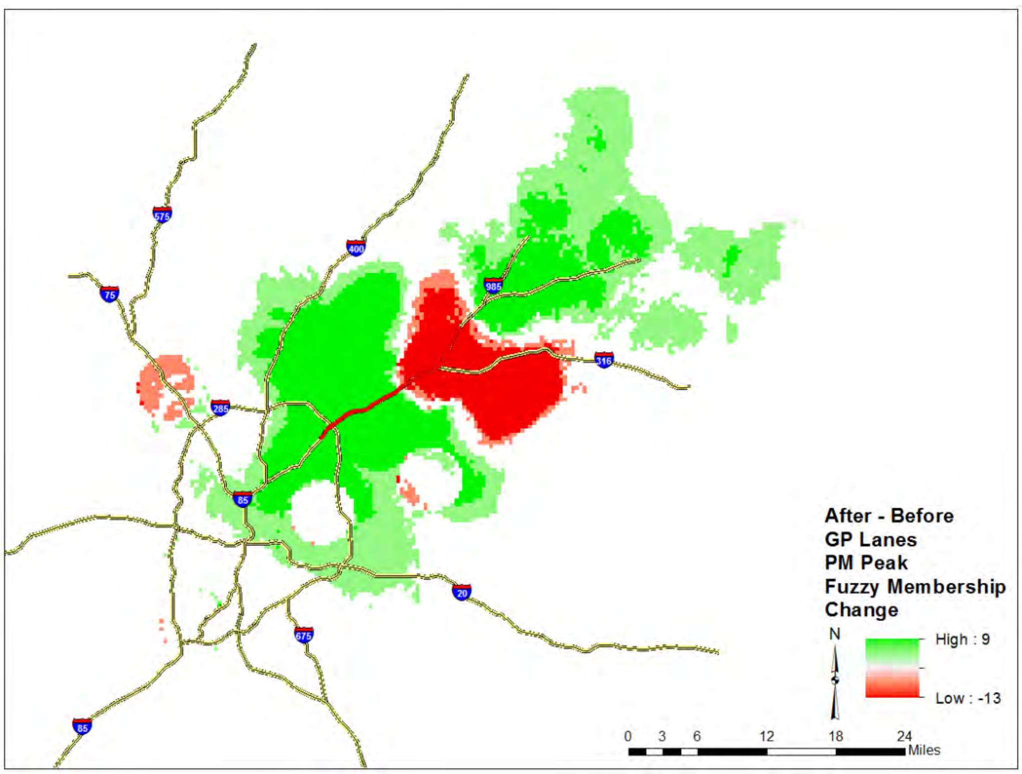


Figure 115 Before and After Conversion GP Lanes Commutershed Change (PM Peak)

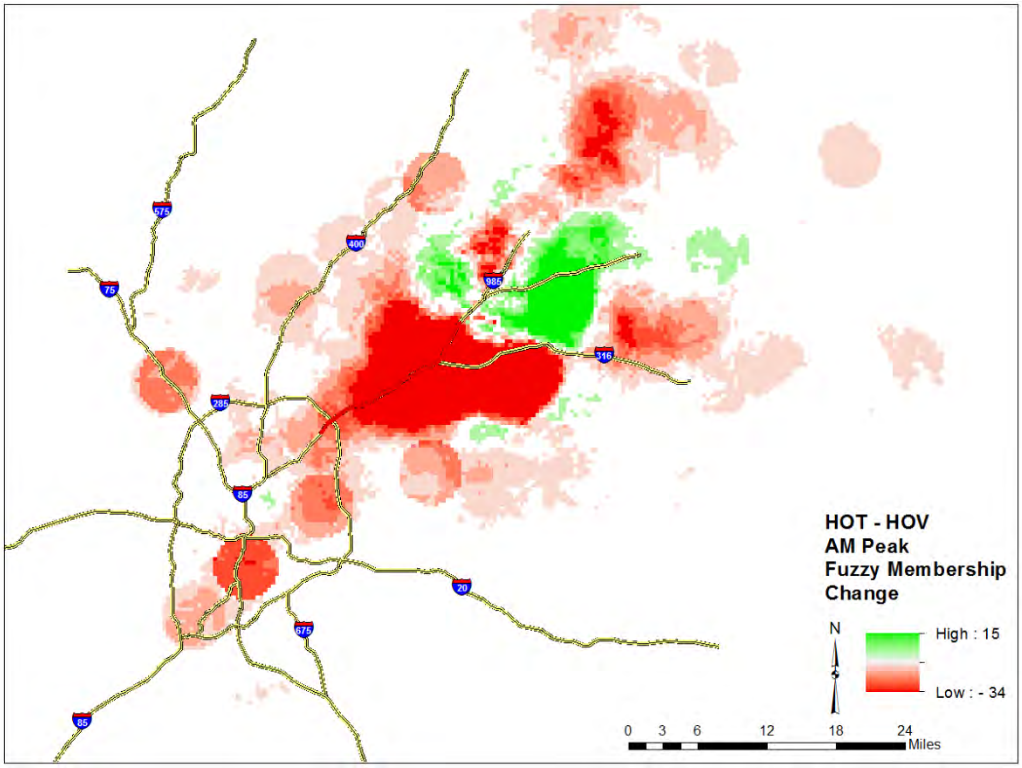


Figure 116 Before and After Conversion ML Lanes Commutershed Change (AM Peak)

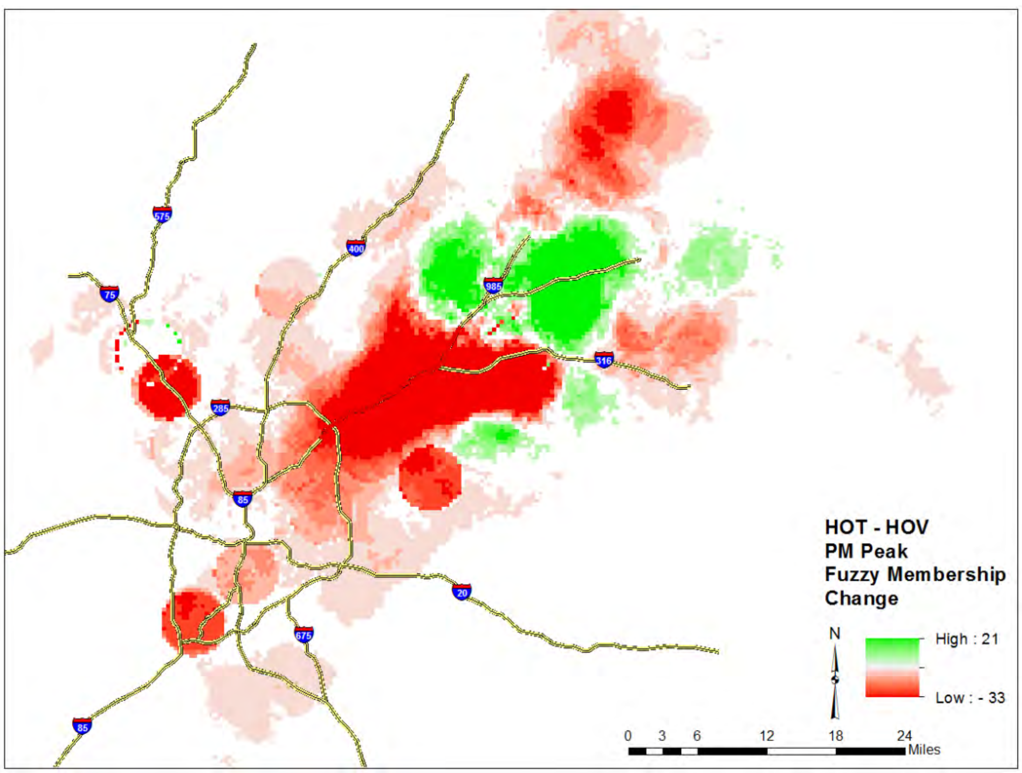


Figure 117 Before and After Conversion ML Lanes Commutershed Change (PM Peak)

REFERENCES

American Community Survey. (2012). *5-Year Summary File Technical Documentation*.

Atlanta Regional Commission. (2011). *Atlanta Regional Travel Survey Final Report.*

Benjamin, J., Sakano, R. (2007). An Analysis of HOT Lanes in North Carolina. *Transportation Research Board 86th Annual Meeting*. No. 07-1246.

Bhat, C., Koppelman, F. (1999). Activity-based Modeling of Travel Demand. *Handbook of transportation Science.*

Braga, A., Bond, B. (2008). Policing crime and disorder hot spots: A Randomized Controlled Trial. *Criminology* 46.3, 577-607.

Burris, M.W., Ungemah, D.H., Mahlawat, M., Pannu, M.S. (2009). Investigating the Impact of Tolls on High-Occupancy-Vehicle Lanes Using Managed Lanes. *Transportation Research Record: Journal of the Transportation Research Board* 2099, 113–122.

Burris, Mark, Nelson, S., Kelly, P., Gupta, P., Cho, Y. (2012). Willingness to Pay for High-Occupancy Toll Lanes. *Transportation Research Record: Journal of the Transportation Research Board* 2297, 47–55.

Burris, Mark, Sadabadi, K., Mattingly, S., Mahlawat, M., Li, J., Rasmidatta, I., Saroosh, A. (2007). Reaction to the Managed Lane Concept by Various Groups of Travelers. *Transportation Research Record: Journal of the Transportation Research Board* 1996, 74–82.

Burris, MW, Hannay, R. (2003). Equity analysis of the Houston, Texas, QuickRide project. *Transportation Research Record: Journal of the Transportation Research Board* 1859, 87–92.

Choo, S., Mokhtarian, P.L. (2004). What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A: Policy and Practice* 38, 201–222.

Cohen, J., Cohen, P. (1975). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.

Colberg, K. (2013). Investigating the Ability of Automated License Plate Recognition Camera Systems to Measure Travel Times in Work Zones. *Georgia Institute of Technology, Master Thesis.*

Collier, T., Goodin, G. (2002). *The Funding and Financing of Managed Lanes Projects* (No. FHWA/TX-03/4190-9).

Collier, T., Goodin, G. (2004). *Managed Lanes: A Cross-Cutting Study* (No. FHWA-HOP-05-037).

D'Ambrosio, K.T. (2011). Methodology for Collecting Vehicle Occupancy Data on Multi-Lane Interstate Highways: A Ga 400 Case Study. *Georgia Institute of Technology, Master Thesis.*

Dahlgren, J. (2002). High-occupancy/toll lanes: where should they be implemented? *Transportation Research Part A*: Policy and Practice 36, 239–255.

Devarasetty, P.C., Burris, Mark, Douglass Shaw, W. (2012). The value of travel time and reliability-evidence from a stated preference survey and actual usage. *Transportation Research Part A: Policy and Practice* 46, 1227–1240.

Dill, J., Weinstein, A. (2007). How to pay for transportation? A survey of public preferences in California. *Transport Policy* 14, 346–356.

Doherty, S. (2009). Emerging Methods and Technologies for Tracking Physical Activity in the Built Environment. *Transport Survey Methods*: *Keeping up with a Changing World*, 153-190.

Douma, F., Zmud, J., Patterson, T. (2005). Pricing Comes to Minnesota: Attitudinal Evaluation of I-394 HOT Lane Project. *Transportation Research Board Annual Meeting*.

ESRI, (2013). ArcGIS Help 10.1. "http://resources.arcgis.com/en/help/"

Fleiss, J., Everitt, B. (1971). Comparing the Marginal Totals of Square Contingency Tables. *British Journal of Mathematical and Statistical Psychology* 24.1, 117-123.

Fuhs, Chuck, Obenberger, J. (2002). Development of High-Occupancy Vehicle Facilities Review of National Trends. *Transportation Research Record: Journal of the Transportation Research Board* 1781, 1–9.

Golob, T., Bunch, D., Brownstone, D. (1997). A Vehicle Use Forecasting Model Based on Revealed and Stated Vehicle Type Choice and Utilisation Data. *Journal of Transport Economics and Policy* , 69-92.

Granell J. (2002). Model Year Distribution and Vehicle Technology Composition of the Onroad Fleet as a Function of Vehicle Registration Data and Site Location Characteristics. *Georgia Institute of Technology, doctoral thesis.*

Greene, E., Peirce, S., Petrella, M. (2012). Longitudinal Household Travel Diary Study : Seattle & Atlanta Urban Partnership Agreement (UPA)/ Congestion Reduction Demonstration (CRD), *Household Travel Survey Symposium*. Dallas, Texas.

Guin, A., Hunter, M., Guensler, R. (2008). Analysis of Reduction in Effective Capacities of High-Occupancy Vehicle Lanes Related to Traffic Behavior. *Transportation Research Record: Journal of the Transportation Research Board* 2065, 47–53.

Hayter, A. (2011). Probability and Statistics for Engineers and Scientists.

Horner, M., Groves, S. (2007). Network Flow-based Strategies for Identifying Rail park-and-ride Facility Locations. *Socio-Economic Planning Sciences. 41*(3), 255-268.

HOV Strategic Implementation Plan Atlanta Region. (2003). "http://www.dot.ga.gov/Projects/studies/Pages/HOV.aspx".

Hultgren, L., Kawada, K. (1999). San Diego's Interstate 15 High-Occupancy/Toll Lane Facility Using Value Pricing. *ITE Journal* 69.6.

Kaufman, L., Rousseeuw, P. (2009). Finding Groups in Data: An Introduction to Cluster Analysis.

Khoeini, S., Rodgers, M.O., Elango, V., Guensler, R. (2012). Sensitivity of Commuters' Demographic Characteristics to License Plate Data Collection Specifications: Case Study of I-85 High-Occupancy Vehicle to High-Occupancy Toll Lanes Conversion in Atlanta, Georgia. *Transportation Research Record: Journal of the Transportation Research Board* 2308, 37–46.

King, D., Manville, M., Shoup, D. (2007). The Political Calculus of Congestion Pricing. *Transport Policy* 14, 111–123.

Kitamura, R, Golob, T., Yamamoto, T., Wu, G. (1999). Accessibility and Auto Use in a Motorized Metropolis.

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

Kressner, J.D., Garrow, L. a. (2012). Lifestyle Segmentation Variables as Predictors of Home-Based Trips for Atlanta, Georgia, Airport. *Transportation Research Record: Journal of the Transportation Research Board* 2266, 20–30.

Kriger, D., Shiu, S., Naylor, S. (2006). NCHRP Synthesis 364: *Estimating toll road demand and revenue.*

Li, J. (2007). Potential Users' Perspectives on Managed Lanes Insights From a Focus Group Study. *Public Works Management & Policy* 12, 416–430.

Litman, T. (2002). Evaluating Transportation Equity. *Editorial board*, *8*(2), 50-65.

Litman, T, Burwell, D. (2006). Issues in sustainable transportation. *International Journal of Global Environmental*. Issues 6.4, 331-347.

Litman, Todd, Brenman, M. (2012). A New Social Equity Agenda for Sustainable Transportation, *Transportation Research Board Annual Meeting.*

Lorenzo, G. Di, Reades, J. (2012). Predicting personal mobility with individual and group travel histories. *Environment and Planning-Part B* 39.5, 838.

MacCullagh, P., Nelder, J. (1989). Generalized Linear Models.

Meyer, M., Miller, E. (2001). Urban Transportation Planning: A Decision-oriented Approach.

Miller, T., Davis, W. (2002). Effect of County-level Income on Vehicle Age Distribution and Emissions. *Transportation Research Record: Journal of the Transportation Research Board* 1815.1, 47-53.

Moran, P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*.

Munnich, L., Loveland, J. (2005). Value Pricing and Public Outreach: Minnesota's lessons learned. *Transportation Research Record: Journal of the Transportation Research Board* 1932.1 (2005): 164-168.

Nelson, J., Zuyeva, L. (2010). A SocioDemographic Analysis of Northeast Atlanta I-85 Peak Period Commuters Likely to be Affected by Implementation of Value Pricing Along the Corridor. *Transportation Research Board Annual Meeting.*

Nelson, J.I., Guensler, R., Li, Hainan (2008). Geographic and Demographic Profiles of Morning Peak-Hour Commuters on Highways in North Metropolitan Atlanta, Georgia. *Transportation Research Record: Journal of the Transportation Research Board* 2067, 26–37.

Order, E. (1994). Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations: Executive Order 12898.

Patil, S., Burris, M., Douglass S. (2011). Travel Using Managed Lanes: An Application of a Stated Choice Model for Houston, Texas. *Transport Policy* 18, 595–603.

Perez, B., Fuhs, C, Gants, C., Giordano, R., Ungemah, D. (2012). *Priced Managed Lane Guide,* (FHWA-HOP-13-007).

Poole, R., Balaker, T. (2005). Virtual Exclusive Busways: Improving Urban Transit While Relieving Congestion. *Reason Foundation*.

Ross, C.L., West, H., Barringer, J., Danner, A., Allen, M., Barrella, E., Doyle, J., Nelson, J.I. (2008). *Congestion Pricing Response: Study for Potential Implementation in the Metropolitan Atlanta Area. "http://www.cqgrd.gatech.edu/sites/files/cqgrd/files/gdot_congestion_pricing_study_full.pdf".*

Safirova, E., Gillingham, K., Nelson, P., Harrington, W. (2003). Are HOT Lanes a Hot Deal ? The Potential Consequences of Converting HOV to HOT Lanes in Northern Virginia Are HOT Lanes a Hot Deal ? The Potential Consequences of Converting HOV to HOT Lanes in Northern Virginia. *Urban Complexities* Issue Brief 03-03.

Schönfelder, S., Li, H, Guensler, R, Ogle, J. (2006). Analysis of commute Atlanta instrumented vehicle GPS data: Destination choice behavior and activity spaces. *ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme*.

Sheskin, D. (2003). Handbook of Parametric and Nonparametric Statistical Procedures.

Smirti, M., Evans, A. (2007). Politics, Public Opinion, and Project Design in California Road Pricing. *Transportation Research Record: Journal of the Transportation Research Board* 1996.1. 41-48.

Sullivan, E. (1998). Evaluating The Impacts of the SR91 Variable-Toll Express Lane Facility. Sacramento, CA.

Supernak, J., Golob, J. (2002). San Diego's Interstate 15 congestion pricing project: Attitudinal, behavioral, and institutional issues. *Transportation Research Record: Journal of the Transportation Research Board* 1812, 78–86.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 227–9.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic geography* 46, 234-240.

Train, K. (1986). Qualitative choice analysis: Theory, Econometrics, and an Application to Automobile Demand.

Ungemah, D. (2007). This land is your land, this land is my land: Addressing equity and fairness in tolling and pricing. *Transportation Research Record: Journal of the Transportation Research Board.* 2013.1 : 13-20.

Wang, P., Hunter, T., Bayen, A.M., Schechtner, K., González, M.C. (2012). Understanding Road Usage Patterns in Urban Areas. *Scientific reports* 2, 1001.

Xu, Y., Zuyeva, L., Kall, D. (2009). Mileage-Based Value Pricing: Phase II Case Study Implications of Commute Atlanta Project. *Transportation Research Board*.

VITA

Sara Khoeini

Sara Khoeini was born in Tehran, Iran. She received a Bachelor of Science in Civil Engineering with a concentration in surveying methods from K.N.T University of Technology, Tehran, Iran in 2007. She earned her master's degree from Clemson University in Transportation and Highway Engineering in 2009 before coming to Georgia Tech to pursue a doctorate in Transportation Systems Engineering. When she is not working on her research, Ms. Khoeini enjoys family and friends gatherings and outdoor activities.