ABSTRACT

The aim of this project was to identify how jobs, housing, transit, demographic makeup, and growth rate -- on the ZIP code level -- predict the average vehicle miles traveled per year per capita (VMT/C) of residents living within that ZIP. This was achieved using ZIP code level housing, demographic, and growth rate data purchased from ERSI and ZIP code level business data procured from the Census. The data were collected for all residential ZIP codes in the United States. Transit availability and use data were obtained at the transit authority level from the American Public Transportation Association. Transit data were collected by urbanized area. Results from this project show that the most influential factors on driving behavior, as ranked by standardized coefficient, are: percent people in families (beta = .16), housing units renter occupied (-.16), total jobs per capita greater than 5 but less than 30 miles from the ZIP centroid (.12), average family size (-.12), percent of people living in group quarters (-.10), population density (-.09), access to bus based transit (-.08), and median household income (.08). Also of note in the findings were the scattered results of the transit variables. Access to bus transit had a strong, suppressive effect on VMT/C, while bus trips per year was insignificant. Train transit availability was insignificant, while train trips per year had a relatively weak, positive effect on VMT/C. Overall the findings align with the existing literature on the subject, however there are enough deviations that it might be wise to question some of the conventional wisdom on how structural features dictate driving behavior. Of the three groupings focused on: transit, housing, and jobs, the effects of housing structure and population were the most influential in predicting VMT. The implications of this finding are that policies aimed at reducing VMT would be most effective if aimed at the structure of housing units and people within those units rather than policies that encourage transit or mixed use development.

Brown University, Box 6070 • telephone: 703.615.2341 • Email: Drew.Raines@gmail.com

Acknowledgements

I would like to thank my advisor, Kurt Teichert, for sticking with me through this project, and over my whole Brown career.

I would also like to thank Sri Nagavarapu and Andrew Aurand for guiding me through the technical aspects of the work, always being quick and enthusiastic with help, and being flexible with the reader/outside advisor situation.

This project would never have been conceived or attempted without the support of my Dad, and never been possible without all the work Philip Moore, Charlie Guthrie, Michael Pierce, Miguel Leno, Larry Majerus, and Gerry Bayer put in from the Carfax end.

Thank you to all the people in the CES and Brown community who lent a hand, or ear, at one point or another: Caroline Carp, Lynn Carlson, Patti Caton, Betsy Barlow, Ross Stackhouse, Yoni Dolgin, and all my other friends who helped in one way or another.

Finally, I would like to thank my Mom. Without her this paper would be one big red squiggly line.

D E V E L O P M E N T S T R U C T U R E S A N D D R I V I N G P A T T E R N S

Using job, housing, transit, demographic, and growth rate data to model and better understand how the physical dimensions of a ZIP code affect the driving behavior of those who live there

BY DREW RAINES

SENIOR THESIS

Submitted in partial fulfillment of the requirements for the Degree of Bachelor of Arts in the Center for Environmental Studies at Brown University

Providence, Rhode Island

May 2010

Brown University, Box 6070 • telephone: 703.615.2341 • Email: Drew.Raines@gmail.com

Table of Contents

List of Figures

I. Introduction

Central Issue

Understanding what forces and factors influence the driving habits of Americans is vital to the ability to manage resources. Driving carries with it a number of externalities; loss of work hours, pollution, global warming, land conversion, water runoff, and highway accidents are just a handful of the effects of our driving culture. The impacts of cars are not all negative; cars also provide what is often the quickest, safest, most convenient means of transportation. If it is society's goal to reduce the negative effects of driving, increasing the cost of vehicle use through such measures as increased gas and or carbon taxes may achieve this end, but will also lower the overall utility [or relative satisfaction] of the population. A less damaging solution, then, might be to reduce the need for vehicle usage. Identifying what factors necessitate increased driving allows the implementation of policies that reduce these factors, and in turn reduce vehicle miles traveled per capita (hereafter "VMT/C"). Reductions in VMT/C using these methods can provide a benefit to individuals, while moving them towards socially desirable driving levels.

Impacts of Vehicle Usage

Environmental

Motor vehicles cause air pollution, water pollution, and other forms of environmental damage.. Emissions from vehicles are believed to cause between $3 and $6 billion in damage to agricultural crops each year (Splash, 1997). Delucchi(1998) estimates that the damages to forests may be in the range of $.2 to $2 billion per year. Environmental degradation through the emission of carbon monoxide, nitrogen dioxide, ozone and particulate matter can lead to human health effects. (McCubbin, 1996) Of these, particulate matter emissions have the largest impact on human health. Vehicle emissions account for between 30 and 40 per cent of all emissions of such pollutants in the United States.

The U.S. transportation system as a whole accounts for two thirds of all petroleum use in the U.S. and causes as much as 30 percent of total US emissions from fossil fuels (Deluchi, 1991, McCulloch 2009). Per capita consumption of oil in the United States is almost 3 gallons per day. Other developed countries average 1.4 gallons, and developing countries average around 0.2 gallons per person per day. (EIA, 2004) Unlike most of the rest of the world, the U.S. uses the majority of this oil for transportation, not heat and power. The myriad effects of global warming caused by this release of green house gas is difficult to quantify, but the portion of the problem caused by vehicles, even restricted to vehicles in the United States, is significant.

Vehicles and the infrastructure that supports them, such as petroleum refineries and roads have effects beyond air pollution as well. The approximately 4 million miles of public roads in the United States have an impact on the land surrounding and underneath them, both in the form of the estimated 1 million vertebrates killed daily by vehicles on those roads, and the less tangible effects of runoff, species migration disruption, and loss of habitat (Forman, 1998).

Importantly, these effects are all externalities of vehicle usage - drivers do not pay for each pound of CO_2 produced or vertebrate killed. The environmental effects of vehicle use are born by society, which, in the case of global warming means future generations across the globe.

Social

The term “automobile dependence” (Newman and Kenworthy, 1989) first coined in 1989, describes the central phenomenon in the much researched field of urban car usage. There has been extensive data collection surrounding the issue due in large part to the social impact. The direct effects of vehicle usage, including unintended injury, costs of maintaining vehicles and infrastructure are quite glaring (Gee, 2004). Other, less obvious effects exist as well. Displacement of city residents, racial segregation due to construction, annoyances due to traffic noise, and time lost to traffic are all major societal costs (Sugrue, 1996).

The lifestyle associated with automobile dependence and the culture of driving also have negative effects on the physical and mental states of humans. Studies of stress levels and children’s mental health levels have shown increased levels of stress and decreased levels of mental health in automobile dependent environments (Gee, 2004). Physically, the automobile dependent culture and the sedentary lifestyle it enables have been linked to obesity levels (Hillamn, 1997)

The costs of vehicle usage are increasing. The United States spent more the $700 billion on oil in 2008, about two thirds of which went to transportation. Especially problematic was that $400 billion of that was oil from other countries. (EIA, 2009) In order to keep up with increasing population and increasing per capita VMT, the United States will have to spend $927 billion dollars just on the construction of new roads to keep up with demand. This does not include the cost of road maintenance, land acquisition, or bridge construction. (Burchell and Mukherji, 2003). Many of these costs are not directly born by the users of automobiles, so the incentive system to maintain this infrastructure is misaligned.

Causes of Vehicle Dependence

<u>Cultural</u>

Why do Americans drive more than the rest of the world? We drive more than Germans, who have a greater number of cars per capita. We drive more than Canadians, who live in a less dense country (U.S. Department of Transportation, 2003) What is it then that pushes Americans to drive more than any other people? There does not seem to be a straight forward answer. There are many structural factors addressed already that contribute to high VMT/C, but the issue is deeper than that. Just looking at structural factors alone cannot fully predict people's behavior. Canadians live at lower densities, so they should drive more. Germans have more cars per person, so they have easier access to driving. Yet we have higher VMT/C rates than either country. What remains then, are the less tangible factors. Could it be that Americans just love their cars more? Or the feeling of freedom that those cars create? Or use their cars as an extension of themselves, a piece of their personality? Some existing research suggests that this could be the case.

Driving behavior is a complicated issue and can not be simplified just to reflect need. Studies show we drive more than we need to because, beyond merely being our primary means of transportation, cars provide emotional comfort and positive feelings. One study (Steg, 2004) showed that especially among frequent drivers and younger male drivers, the value of such symbolic and emotional benefits of driving were often primary even over the utility such driving provides. It must be assumed then that just looking at the distances between people's places of work and dwellings is insufficient in order to fully comprehend the causes of VMT.

The speed, flexibility, and convenience of vehicular travel are major factors in the choice many people make to drive over other alternative forms of transportation. However, just from looking at how cars are advertised it is clear utility is not the driving concern; car advertisements use feelings of power, fun, superiority, and sexiness as the primary drivers to sell vehicles. While things such as storage space, safety and gas mileage are surely used in advertising, social status and self esteem are highlighted more aggressively than logic. Other studies such as Stradling et al. (2000) found that those who valued being independent and who reported deriving some sense of personal identity through their cars were less likely to reduce car use. It is thought that there are three basic values of a car: the obvious utility it enables, the social platform it provides for self expression and social position, and the emotional effect driving has on the mood of drivers.

Traditionally, economic models account for travel time as a cost. The assumption is that time spent traveling to and from activities is a major force keeping people constrained. However, human desire is complicated, and sometimes car use is a neutral or even positive element, and may be desired even for its own sake (Mokhtarian, 2001). Part of the reason it is so hard to get people to switch from driving to public transportation even when the utility value generated by the two appear to be equal is the difference in the perception of these two modes of travel. A study (Steg, 2004) reported that even infrequent car drivers have strong positive feelings towards traveling by car, while only a minority of people have strong positive feeling about public transit. While this study was conducted in Denmark, some of the findings are likely to be relevant to the United States.

Further complicating the matter, some studies show that self selection may be a large, and largely ignored factor in predicting vehicle usage. (Handy et al, 2005) They argue that individuals who do not wish to drive opt to live in neighborhoods that are conducive to a more limited driving lifestyle. In this case, the neighborhood does not cause them to drive less, rather their choice of neighborhoods reflects their desire to drive less. However, that people chose an area where they can drive less speaks to the physical properties of the area that allow it to be conducive to driving less. A high density area, then, may be conducive to lower driving habits, and therefore have lower-than-average VMT/C. Some part of this lower-than-average VMT/C might be due to the fact that people who want to drive less live there, but it seems likely that the physical structure of the area is responsible for VMT as well. The physical characteristics of the geographic area enable the behavioral preferences of the residents.

Structural

The form and structure of our built environment have a direct impact on the amount people drive. If one lives farther from one's place of work, one needs to drive farther to get there. If one lives farther from the businesses that provide goods and services, such as banks, grocery stores, clothing stores, and restaurants one needs to drive farther to complete necessary errands. The basic principles and trends, therefore, are relatively straight forward. Quantifying the effects of structure is a much more complicated matter. There are an almost infinite number of structural characteristics of an area that can alter VMT. These characteristics range from the obvious - number of people per square mile - to the minute - the existence of painted crosswalks. Complicating matters further, different cultural conditions affect the relative importance of all these characteristics on VMT. Truly understanding and being able to predict VMT is tantamount to being able to predict any human behavior; to accurately be able to predict how an individual will behave, one needs to know every stimulus they experience. As understanding and controlling for every input is impossible, the best one can do is to gather as much data as possible about an area, and hope to get a relatively accurate picture.

Sprawl

Sprawl is one of the major issues that is always addressed in conversations about VMT. While sprawl is not the whole picture, it is an important example of how the structure of development affects VMT, and how our current development practices are requiring more and more vehicle dependence. Sprawl is hard to define accurately, because what constitutes sprawl is different on the micro level than it is on the macro level. Most basically, urban sprawl is "low density, automobile dependent development beyond the edge of service and employment areas" (US EPA, 2003) It is sometimes measured by use of the U-index which is a measure of the total area that is covered by either urban or agricultural lands. The U-index is not in widespread use in this area of study, because the definition of sprawl is too complicated to be narrowed down to one term. Sprawl is difficult to pin down because it comes in so many varieties, with varying characteristics. Sprawl does not just mean any construction on a greenfield, or previously undeveloped land. Sprawl is development merely based on that rapidly expands into greenfields beyond the scope of organized infrastructure expansions. Inherent in the concept of sprawl is the idea that it outpaces the infrastructure expansion of the central city. Not all low density areas are sprawl, so any classification that attempts to draw lines just based off of density misses much of the point. At the macro level you need to look at surrounding areas, geography, and construction timelines to really identify sprawl. On the ground, at the micro level, factors such as municipal water, school locations, and business parks can help an observer identify areas of sprawl.

Governmental regulations have affected the occurrence of sprawl and sprawl-like devolopments In 1944, after WWII the introduction of the Servicemen's Readjustment Act, otherwise known as the GI bill, made possible the picturesque stability of suburbs for all those whose peace was torn by the war. This Act provided loan guarantees for home mortgages for WWII vets, making the American dream a reality for those for whom it would have otherwise have been out of reach. (U.S. Department of Interior) Along with this public spending on the infrastructure to reach the suburbs, support buying suburban homes heavily affected settlement patterns after WWII, turning us from an urban country into a suburban one.

US development practices encouraged sprawl through strengthening and adding to road networks, in particular the interstate highways system. It is only possible to live in spread out communities distant from commercial and industrial centers if there exists a stable and convenient transportation system for getting to those places of work and commerce. The federal money spent subsidizing the construction of highways and local roads broke down what would have otherwise been a larger cost barrier to living in the newly forming suburbs. If the costs of road construction had been borne by those who used them, i.e., those living in the suburbs and requiring the highways to get to their places of work, then it is quite possible that settlement behavior would have been significantly different. If all new roads were paid for by tolls, a commuting lifestyle might have been less financially attractive and therefore less popular.

A study (Nasser, 2001) found that the aridity of a region is a major limiting factor for sprawl. If development cannot dig wells for its water, it cannot leapfrog open spaces, but must instead expand in a tighter manner and stay close to the municipal waters lines necessary for life. Similarly, if a city is boxed in by oceans or mountains it will tend to grow more compactly, or at least it will after it reaches the limits of its usable land. Finally, the study found that a lack of regional governmental planning leads to more sprawl: the study noted that the Northeast and Midwest tend to have both lots of sprawl and decentralized regulations.

Development Planning

There are number of movements that espouse patterns of development they believe result in both lower VMT/C and a higher standard of living. Two of the best known of these organizations are Smart Growth America, and New Urbanism. The core ideology of both of these movements is that denser, more mixed, areas with access to public transportation have a higher quality of life, in part because of reduced dependence on driving. (Smart Growth, 2010; New Urbanism, 2010)

There, however, exists very little hard evidence to quantify and confirm these central tenants. Studies that show exactly how much density, for example effects VMT/C does not exist, at least not at a national level. The goal or this project, then, is to add hard numbers to some of the often repeated rules of thumb, and add credibility to movements such as Smart Growth and New Urbanism.

II. Literature Review

Major Types of Existing Studies on VMT

There is a large body of literature in existence on the subject of driving habits, sprawl, and development patterns. Of particular interest to this study are those researchers who have attempted, through a variety of methods, to determine what affects the amount people need to drive.

There are three major types of existing studies on driving behavior. The first uses traditional transportation models to predict differences in VMT between a given, traditional area, and that of a hypothetical alternate. This type is mainly used to explore sprawling areas, and determine how a non-typical structure could change the average VMT. The second type of study looks at the micro level to examine the differences between small areas, and inquires into the relative importance of urban form factors, such as sidewalk shading or mixed use areas, to vehicle usage. A final, similar, study tends to look more at the macro level and uses aggregated data to compare average travel characteristics of one area to that of another location.

Each of these methods asks a slightly different question and returns different results. The first and third methods tend to show that suburbs have significantly higher average VMT levels than do more dense areas, or those possessing traditional design characteristics. The second method is more mixed in these findings. (Handy, 1996) hypothesizes that this difference is due to the fact that more macro studies lump too many characteristics together, and that actually the picture is far more complicated than those simplistic models allow.

<u>Transportation Models</u>

Many of the transportation model based studies reviewed focus on comparisons between sprawl and other types of development. Robert Burchell has been lead writer on many of these papers, including “The Cost of Sprawl” (Burchell, 2001) and “The Incidence and Cost of Sprawl in the United States” (Burchell, 2003). These papers have attempted to quantify exactly how sprawling new developments are becoming, what it is that defines sprawl, and ultimately what the social, environmental, and economic costs of sprawl are for the United States.

<u>Overview Studies</u>

One of the recent major studies in this area was conducted by Steven Polzin and published in 2006. The study uses National Highway Transportation Survey (NHTS) data to study the factors that influence travel behavior. They break these factors down into three major categories: socio-economic, land use, and transportation conditions. Of the three, the study focuses primarily on the effects of socio-economic factors on vehicle miles traveled. The study takes an historical look, tracking trends since 1969 when the first NHTS survey occurred. One of the primary finding of the study was that the US has reached a large junction in trends – that many of the factors that influence VMT have started to settle after periods of consistent increase or decrease.

Since 1977, daily travel has grown 151%, total VMT has grown 90%, while population has only grown 30%. The major socio-economic factors that the study looked at were the age profile of the population, availability of automobiles, household size, license rates, ride share, walking, transit use, the male/female ratio , income, and land use patterns. Socio-economic factors control the demand for activities and the ability to afford such travel activities. Land use patterns and the availability of transportation options impact through which method of transportation the demand for travel manifests. A similar study conducted in 2009 focusing on the effects of shifting demographics in the United States found similar results using very similar methodologies (Contrino, 2009).

Area-Specific Studies

Due to the availability of VMT data, and the general elevated interest level in the geographic retion, the San Francisco bay area has been the subject of multiple area specific studies. Among these, "Travel demand and the 3Ds: Density, Diversity, and Design" (Cervero, 1997) was the most influential on this project. The paper highlighted three overarching structural characteristics that affect VMT: density, diversity, and design. The researchers looked at the effects of a number of variables on the number of trips, the distance of trips, and the transportation mode choice of trips. Density, diversity and design were chosen by the experimenters as the meta variables because they felt that all of the individual variables that they chose to study fell under one of those categories.

Basis for Variables Used

Population

Because personal activity is closely related to travel level, and personal activity is also closely related to the age of an individual, the population profiles for an area are strong indicators of VMT. VMT levels grow with age, because of the impact of children - who do not drive - until somewhere around middle age. Work related travel also peaks around middle age. Once this threshold is passed, trends tend to point downward, as the demand from children and jobs decreases or drops off entirely. Polzin also found that in recent data sets, VMT has increased across the board, and the rates of decline after middle age are slower than they once were, as the elderly are wealthier and hold onto their jobs longer.

Household Size

Average household size has been decreasing since somewhere around 1900 (Polzin, 2006). However that decrease seems to be plateauing. Larger households appear to allow a lower VMT per household, as trips apparently are centralized. One interesting finding was that two person households more than double the VMT of one person households, breaking this apparent economy of scale. One reason for this might be attributed to the makeup of one person households, which tend to be heavily comprised of the elderly and more solitary individuals.

Density

There have been a number of studies since the 1990's that link urban intensity and automobile dependence. (Newman 2006) found that automobile dependence decreases with increased density. He observes a threshold at around seven dwelling per acre, which appears to allow for efficient public transportation and increased walkability. This finding rests on the shoulders of an earlier paper by the same author (Newman and Kenworthy, 1999), which shows that when urban densities exceeded thirty persons per hectare, automobile use declines rapidly. Even as early as 1977 a study (Pushkarev and Zupan, 1977) found that foot travel increased in more traditionally arranged communities as compared to those with more modern suburban subdivisions.

Racial Diversity

Though highly correlated with income, the effects of diversity and shifting diversity patterns on VMT provide a unique insight. Many more minority households depend on alternative forms of transportation, a fact driven by lower vehicle ownership rates of 1.38, 1.74, and 1.69 vehicles per household for Black, Asian, and Hispanic households respectively. (Contrino, 2009). In comparison the average White household has 1.99 vehicles.

Job Locations

Ewing et al., (1994) found that mixed use neighborhoods induce shorter trips. He concluded that much of the VMT lowering effect attributed to higher densities could actually be the effect of mixed use increases, which tend to go hand in hand with increases in density. Diversity was measured using a very complicated grid system: using hectare square grids, diversity points were allocated for containing or being close to a number of goods and services. (Newman 2006) measured urban intensity using job density as well as population density. His findings regarding job density also show an inverse relationship between density and VMT/C.

Income

Income drives VMT in two ways: it enables persons to afford travel costs, and drives demand for travel through activities such as entertainment and shopping. However, it appears that the trends are towards lower marginal operation costs for vehicles, meaning that it's not the cost of travel that enables higher VMT at higher incomes, but the ability to afford activities that necessitate travel. (Polzin, 2006) Income is also heavily correlated with a number of other factors which influence VMT, such as age, vehicle availability, or female labor force participation. It is important, then, when looking at income effects to view it in context with its tertiary effects. There is no consensus on the effects of income on VMT, and authors such as Newman, believe that it is a non-issue in regard to driving.

Transit and Carpooling

The data on transit use is incomplete, but it appears that commuter transit use has dropped from around 8.9% in 1970 to 4.7% in 2000 and from 3.4% to 1.56% for all travel. “Even if all transit disappeared – a virtually impossible outcome – the impact on aggregate VMT would be limited to less than two percent.” (Polzin, 2006) Carpooling is experiencing similar declines, likely based on greater automobile availability, and more dispersed driving and working patterns. As carpooling is already quite low, and much of what there is comes from child chauffeuring, it seems unlikely it will drop much lower in the near future. (Polzin, 2006)

Unique Additions to the State of Knowledge

Fit of the Study with Existing Literature

As the majority of previous studies fall into the three categories described earlier, there exists a gap of knowledge as to the effects of small-scale structural units aggregated across the whole country. In particular, this study fits between the overview studies and the location specific studies; providing location specific results across all 50 states.

Most macro studies on this subject are intended to study nation-wide trends in automobile usage, using aggregate, survey based data sources for VMT/C as well as other variables. These studies, such as (Polzin, 2006) and (Contrino, 2009), treat the country as a whole, or a small number of large pieces. They are very useful in tracking how wide demographic shifts across the country are have affected (or are predicted to affect) such factors as VMT/C, however they are able to tell very little about how these trends interact with and affect VMT/C on a geographic level more similar to a neighborhood than a Metropolitan Statistical Area. One of the hurdles for this type of study being applicable down at smaller geographic areas is the lack of universal and geographically accurate VMT/C data. VMT/C data used in previous studies tends to come from ether small scale studies, such as the smog study in San Francisco which Cervero and Kockelman 1997 used for their research, or larger survey based studies that measure vehicle usage via surveys of driver behavior. The most prominent of these surveys is the National Highway Transportation Survey.

On the other hand, the other variety of VMT modeling uses constrained geographic areas, for example (Pushkarev and Zupan, 1977) or (Cervero, 1997), for which VMT/C and the variables hypothesized to affect it can be meticulously measured. Studying only a small geographic area allows unique data sets such as the San Francisco smog study to be used. Studies that cover only a small geographic area have the added benefit of allowing for the creation of unique variables. Ewing was able to develop a complicated and labor intensive grid to measure mixed use in the Bay area (Ewing et al, 1994), while such a grid would be impractical to apply across the whole country. The disadvantage of location specific studies is that they are more difficult to generalize to other areas.

There are many things unique to, for example, San Francisco that affect the city's average VMT/C. It is impossible to capture all these factors, so any extrapolation of results outside of the city runs the danger of drawing false conclusions. It is dangerous, then, to apply the findings of area specific studies as general rules.

The aim of this study is to fill some of the gaps between these poles. The study examines jobs, housing, transit, demographic makeup, and growth rate, but applies them to VMT predictions at the relatively small unit of a ZIP code. Using such robust and geographically deep data sets, it is possible to make generalized rules of the effects on VMT at the ZIP code level. Most of the data used by this project are widely available to researchers. A barrier to this type of study to date has been the lack of a source of VMT data with enough depth to calculate VMT/C at the ZIP code level, and enough breadth to allow the study site to cover the whole United States.

This study was able to overcome this historical barrier through the use of a novel method of measuring VMT/C. Working with the company CARFAX, this study used odometer readings from over 50 million unique vehicles across the country to create robust VMT averages for all residential ZIP codes. The VMT data culled from odometer readings provided by CARFAX enabled this type of research for the first time.

III. Research Methods

Scope of Coverage

The aim of this project is to quantify the total and relative effects a number of demographic and structural factors have on average VMT per capita. The study area is all 50 states and the District of Columbia. A large study area was chosen for this study because the aim is to produce generalized rules regarding VMT, not examine geographic-specific effects. The aim of the project is to answer questions such as "In the United States, what is the effect of access to bus transit on VMT/C -- controlling for jobs, demographics, housing, and growth rate?" not "Does the construction of additional rail lines in Pawtucket RI decrease VMT per capita?" The essential difference in these questions is that one seeks to create a rule that can be applied with relative accuracy across the United States, while the other applies more specifically to a set of geographic conditions.

Unit of Analysis.

The unit of analysis used for this study was the ZIP code. ZIP codes are rarely used as a geographic unit because of their inconsistencies in size, shape, and population, as well as their frequency of change. Most data that are collected on the zip code level, especially when collected by the census, are actually collected at the block level. Block's are much more stable than ZIP codes, and can relatively easily be assigned to residential ZIP codes using the latitude/longitude of the block centroid, and the boundaries of the ZIP code. This method only works when the ZIP codes have an assigned boundary polygon. Post Office ZIP codes and single address ZIP codes then cannot be assigned demographic properties, and these demographics are incorporated with the data for the enclosing residential ZIP code. The ZIP code boundaries used in this project were current as of November 2008.

For this project ZIP codes were chosen as the geographic unit despite their drawbacks. This decision was driven by three primary factors: the odometer readings which were used to calculate VMT/C were tied to ZIP codes, many commercial businesses produce useful data at the ZIP code level, and ZIP codes are a small enough geographic unit that they tend to be internally homogeneous.

Even though ZIP codes are not a stable geographic unit, they are popular with marketers and other businesses that rely on geographic units to track customers. This is partially due to their connection to mailing addresses, as that is the way the businesses often reach their customers. This trend is also driven by the fact that ZIP codes are geographically small and relatively internally homogeneous. Companies such as ESRI, then, produce much of their data at the ZIP code level, to serve their corporate clients. With the census 2000 data too old to use with the 2009 VMT/C data, commercial data sources, such as ESRI was the best source for the data needed.

The aim of this project was to be able to model and predict vehicle usage at as small a geographic unit as possible. ZIP codes are small enough that they usually can be cleanly classified as urban, suburban, or rural. They also tend to be homogenous for other factors as well, mainly due to their small size.

Sampling Period

The data for this project are from the years 2007 to 2009. Some flexibility was accepted to allow for the inclusion of the Census business patterns data, for which the survey took place in 2007.

It is important for all the data to come from a relatively small timeframe because the methodology of the project was to take a snapshot of America and derive the rules about VMT/C from the patterns found in that snapshot. Alternative methodologies explored to answer similar questions proposed to use variations over time of demographic characteristics. These methodologies were not used for this study due to the increased complexity, and the lack of sufficient data with a temporal attribute

A number of the variables used in the regression do include some measure of time. The 2000-2009 annual compound growth rates of both population and median household income were included in the study. They are included because they are hypothesized to be indicators of new development or growth, and so possible signals for sprawl -conducive conditions.

CARFAX Data

CARFAX Inc is a private company that provides vehicle history reports on used cars. In order to provide these reports, CARFAX has collected and maintained a huge database of information on all cars. Whenever a car gets inspected, gets repaired, or changes owners, data is produced that is collected by CARFAX. One of the pieces of information that is collected is odometer readings. The odometer is a gage that all cars have that records the number of miles that vehicle has driven since it was produced. Using the odometer readings it is possible to calculate the average distance that a given car has driven in a given year. This calculation is only possible if there exist enough and recent enough data points for the vehicle. Using the database assembled by CARFAX, it is possible to calculate the average miles driven for more than 50% of all cars.

Odometer readings were selected to reflect the vehicles this project is interested in measuring: privately owned passenger cars on the road in 2008. Odometer readings from over 50 million unique vehicles were used for this study. Miles traveled for each car for the sample year of 2008 was calculated using miles driven between the most recent two (or more where available) odometer readings, and the amount of time between the readings. The average miles driven per year for each car was averaged by ZIP code. This number was then multiplied by the number of active cars in each ZIP code -- a data set lent to this project by POLK -- to determine the approximate total number of miles driven by all cars in each ZIP. Finally, this pooled mileage was divided by the population of the ZIP code, resulting in the final variable vehicle miles traveled per person per ZIP.

Appendix A (Detailed Methodology of VMT/C Derivation) includes additional details on the process of creating the DV.

ESRI Data

For the reasons mentioned earlier, this project was conducted using data from as close to 2009 as possible. The great disadvantage of this methodology is that the most complete source for all of this type of data is the decennial census. The most recent decennial census was in 2000, therefore much of the data used was based on statistics true in 2000, and projected forward to 2009.

Projections are derived from current events and past trends. While there may not be direct survey data at an individual level for 2009, there is a lot of information about trends over the nine years since 2000, as well as more recent data taken at larger geographic levels. This type of data can inform the development of projections.

There is undeniably some inherent danger to using projected data based on older samples. 2009 is a difficult year to use as a sample year, because the interval is as large as possible from the last 10 year census, and any deviation of the projected trends from the actual events are amplified over time. Fortunately there are a number of other data sources put out by the census bureau and other groups which can be used to refine the projections as the years progress, so they are more accurate than pure projections. The risk of using projections is also amplified with smaller size areas, so using ZIP codes as the basic geographic unit brought additional challenges. Finding the most reliable source for data, was one of the most important aspects of this project.

Appendix B (ESRI Company Description) contains additional information about ESRI, the source of the ESRI variables.

Census Data

County Business Patterns is an annual data set that provides information on the number, size, and industry of most business in the United States. It is survey-based and collected annually. For this project, County Business Patterns 2007 were used, as they represented the most current available data set. The data are available in geographic units as small as ZIP codes. Industry is broken down by 20 categories which correspond to the North American Industry Classification System (NACIS). Excluded from the series are self employed individuals, employees of private households, railroad employees, agricultural producing employees, and most government employees. While the lack of these categories is unfortunate, they represent a small portion of the whole, so their absence can be accommodated.

As the location and type of business in neighboring ZIP codes was hypothesized to have an effect on VMT/C, a methodology was created to measure this effect. Rings were created to sum the number and type of business at different distances from the ZIP code in question. Rings were made for 0-5 miles from the home ZIP, 5-10 miles, 10-20 miles, and 20-30 miles. It was also modeled using industries separated and grouped together. The methodologies and the advantages and disadvantages are listed in Table 1.

Table 1: Exploration of Methods for Jobs Variables

ANALYSIS METHOD	ADVANTAGES	DRAWBACKS
Number and type of workplaces by ZIP code	A good method of exploring how 'mixed use' an area is. The greater the number and the greater the variety of businesses in a ZIP the more mixed use the ZIP can be considered	Half of the industry grouping categories are insignificant, even when run in isolation. Working patterns often cross ZIP lines, so misses many factors that might be influencing VMT
Workers by employment industry by ZIP code	Can be used in conjunction with workplaces per ZIP to assess how many workers commute outside the ZIP.	It is an indirect method of looking at business locations. Not worth inclusion because of collinearity issues.
Jobs per worker by industry for rings 0-5, 5-10, 10-20, and 20-30 miles from each ZIP	The most comprehensive method of analysis. Can tell about how both the spatial and industry break down of businesses affects VMT. Highest R squared value	Many of the variables are insignificant. There are also extensive issues with multicollinearity
Jobs per worker by industry for rings 0-5 and 5-30 miles from each zip	Compromises some of the spatial details of the multiple ring method, but all variables are significant when used in isolation. High R squared value.	When other variables are added to the regression, patterns of the coefficients break down, significance is lost, and interpretation becomes difficult.
Jobs per worker total for rings 0-5 and 5-30 miles from each zip	Analysis of results is simple. One of the variables retains significance even when all other factors are included in the model. Low multicollinearity.	Cannot provide information about the effect of business type.

Appendix C (Census Job Data Creation) contains additional details about the assumptions made in the transformations, and the creation of the job rings.

APTA Transit Data

Access to bus and train transit statistics were taken from the Public Transportation Fact Book. The data tables used were in Appendix B of the Fact Book. Data for Appendix B are taken from the Federal Transit Administration's National Transit Database (NTD) and include only agencies reporting to the NTD. The NTD collects data via an internet based reporting system on over 660 transit providers in the United States. The data collected by the NTD are a nearly complete picture of transit in the Unites States, as recipients or beneficiaries of grants from the Federal Transit Administration (FTA) under the Urbanized Area Formula Program or Other than Urbanized Area (Rural) Formula Program are required by statute to submit data to the NTD.

Data are available for major modes of public transit, such as bus, commuter rail, heavy rail, light rail, and trolleybus. Data are also available for more niche forms of public transportation such as automated guideways, cable cars, ferry boats, inclined planes, monorails, paratransit, and vanpools.

For the purposes of this project, all bus-like transit was grouped into the category of bus. These modes were bus and trolleybus. All train-based transit -- commuter rail, heavy rail, and light rail -- were similarly grouped under the heading of ‘train.’

Data are also summed and ranked for urbanized areas. In order to calculate trips per capita, the total number of unlinked passenger trips for each urbanized area was divided by the total population of that urbanized area. As the analysis for this project was performed at the ZIP code level, in order to match the transit data to the existing data sets, all ZIP codes that fall within an urbanized area were assigned the transit values associated with that urban area.

Statistical Methods in Use

Selection of the Variable Set

For each variable considered (Appendix D), the descriptive statistics were run, the variable was put into a linear regression in isolation with VMT/C, and correlations with each other variable being considered were run. This step measured the distribution of each variable, the isolated predictive power of each variable on VMT/C, and the correlation each variable held with VMT.

A second set of regressions was run on variables that showed significant results. These variables were run in regressions against VMT/C, with median household income (from here on 'income') and/or density included as well. As income and population density ('density') were hypothesized to be highly correlated with many other variables, this second round of regressions was intended to weed out variables for which the effect on VMT was closely related to the income or density, and not unique to the variable.

Table 2: Alternate Means of Measurement

Bold = method used

MEASURE	POSSIBLE MEANS OF MEASUREMENT	TABLE LOCATION
Transit	**Bus**	Appendix G
	Train	
	Other	
	Transit available	
	Unlinked passenger trips per person per year	
Jobs	Jobs rings 0-5, 5-10, 10-20, and 20-30 miles	Appendix H
	Job ring 0-5, 5-30 miles	
	Job type broken by industry group	
	All jobs total	
Racial diversity	**Diversity index**	Appendix I
	% population by ethnic group	

The correlation table was used to explore which variables were closely related, and therefore should be removed from the system to minimize multicollinearity. The final stage of weeding out variables for the regression system was a linear regression that contained all variables considered significant after the first two rounds. This regression was run multiple times to determined the optimum mix of variables. It was run as a stepwise regression, in which variables are added one at a time in order of influence (as determined by standardized coefficient), and variables that added no significance when entered were removed. Stepwise regressions help determine which variables are most important to include, and which have less significant effects or effects too closely tied to other variables.

Some measures had multiple means of measurement. (Table 2: Alternate Means of Measurement) In these cases, extra regressions were run in order to pick the optimal combination of variables.

After the final list of variables had been determined, the remaining variables were grouped into five categories as listed in Table 3. The groups were created to simplify the next stage of the regression.

Table 3: Grouped Variables

GROUP	VARIABLES					
Jobs	Unemployment Rate	Total Jobs per Cap > 5 <= 30 miles	Total Jobs per Cap <= 5 miles			
Demographics	% of Pop 15-64	% Pop 25+ with a Bachelor's Degree or Higher	Median Household Income	[racial] Diversity Index	% Male	
Housing	% of people in families	% of people group quarters	Average Family Size	Population Density	% Housing Units Renter Occupied	% Housing Units Vacant

GROUP	VARIABLES					
Transit	Train trips per Year (un-linked) (by city)	Train Transit Available (by city)	Bus trips per Year (un-linked) (by city)	Bus Transit Available (by city)		
Growth rate	2000-2009 Population: Annual Compound Growth Rate	2000-2009 Median Household In-come: Annual Compound Growth Rate				

Stepwise regressions were then run to help determine how the predictive effects of the variables interacted. The groups 'jobs,' 'transit,' and 'housing' were selected as the three primary groupings. This selection was made because these groups represented the variables of most interest to the study, while 'demographics' and 'growth rate' groupings were included not because they were of interest as primary drivers, but because it was desirable to control the other groups for their demographic and growth rate patterns. Three stepwise regressions were run on the data, each one starting with a different primary group, and adding the other four groups sequentially to observe how the addition of additional variables both affected the predictive value of the variables in the model already, and of the model as a whole. The sequence of additions for these three regressions is outlined in Table 4. Descriptive statistics (Appendix E: Descriptive Statistics for Finals Variables) and correlations (Appendix F: Correlations for Final Variables) were also run on these final variables

Table 4: Regression Order

1	JOB	TRANSIT	HOUSING
2	Demographics	Housing	Demographics
3	Housing	Demographics	Job
4	Transit	Job	Transit
5	Growth rate	Growth rate	Growth rate

Descriptive Statistics (Appendix E)

Mean was used as the measure of central tendency, and the standard deviation from the mean was calculated to understand the spread of the data.

Correlation (Appendix F)

Correlation is the departure of any two or more random variables from independence. This project used Pearson's correlation, which only measures departures from randomness that fall into a linear pattern. It is also most effective when examining only two variables. A Pearson's correlation coefficient of 1 means that the variables are completely linearly related; a coefficient of 0 means that there is no linear relation. Correlations were not used heavily for this project and did not contribute much to the results. The Pearson correlation test was exclusively used in conjunction with regressions on only two variables.

Significance (Tables 6-9)

Significance is the probability of seeing a result as extreme as the one produced, in a collection of random data in which the variable had no effect. A sig. of .05 or less is the point at which for this analysis the null hypothesis has been rejected. With sig. value of .05 there is only a 5% chance that results produced are the same as those in a random distribution, it is therefore possible to say with a 95% probability of being correct that the variable in question is having some effect. Due to the large data sets and the way in which independent variables were chosen for this project, almost all variables used had significance of .000, which indicates at least a 99.99% probability of being correct that the variable in question is having some effect. Cases that deviate from this pattern are expressly noted.

R squared (Tables 6-9)

R squared represents the fraction of the variation in the dependent variable that is predicted by the combination of all independent variables. In the regressions then, with only a single independent variable, the R squared is the same as the square of the correlation.

Unstandardized Coefficients (B) (Tables 6-9)

In a linear regression, the size of the coefficient for each independent variable tells the size of the effect that variable has on the dependent variable, and the sign on the coefficient (positive or negative) tells the direction of the effect. In the regressions run with a single independent variable and the dependent variable, the coefficient tells you how much the dependent variable is expected to increase (if the coefficient is positive) or decrease (if the coefficient is negative) when that independent variable increases by one. In regression with multiple independent variables, the coefficient tells you how much the dependent variable is expected to increase when that independent variable increases by one, holding all the other independent variables constant.

Standardized Coefficients (Beta) (Tables 6-9)

Standardized coefficients are the estimates resulting from an analysis performed on variables that have been standardized so that they have variances of 1. These estimates were used in this project to answer the question of the relative effects of the independent variables on the dependent variable of VMT per capita. This extra step was needed to allow direct comparison between the independent variables because the variables were measured using different units of measurement. For some IV's, like the diversity index, the data were on a scale of 0-100, while others, such as income, were measured in dollars.

All variables (independent and dependent) were standardized by subtracting the mean and dividing by the standard deviation. The standardized regression coefficients, then, represent the change in terms of standard deviations in the dependent variable that result from a change of one standard deviation in an independent variable.

Multicollinearity

Multicollinearity refers to a situation in which two or more explanatory variables in a multiple regression model are highly correlated. In a situation of perfect multicollinearity, the correlation between two independent variables is equal to 1 or -1. In any single variable regression, therefore, by necessity collinearity tolerance is equal to 1, because 100% of the variation in the model is unique to population density, there exist no other variables in the model.

Multicollinearity can become an issue for this project in the multiple-independent-variable regressions. It needs to be monitored particularly carefully when two or more variables used in the same regressions are known to be highly correlated, such as average family and average household size. SPSS produces a collinearity output variable for each independent variable. This then can be used to understand which variables are too closely related to both be included.

The final regression for this project trimmed most of these variables out (for example leaving average family size, but removing average household size).

Multicollinearity does not reduce the predictive power or reliability of the model as a whole; it only affects calculations regarding individual independent variables. Even with highly correlated IV's, a model can indicate how well the entire set of IV's predicts the dependent variable, but it may not give valid results about any individual IV's power to predict.

Software

The majority of the analysis for this project was performed on SPSS. SPSS is widely believed to be among the most intuitive statistical packages available. The availability of SPSS both at the CARFAX office and on the Brown University campus additionally catapulted it above the alternatives. Data organization and transformation was performed using both SPSS and Microsoft Excel. Additionally, the calculation of 5, 10, 20, and 30 mile rings was performed using the ESRI software ArcView.

Data Preparation

To allow comparison between variables, all variables were transformed in such a way that each variable had a value for each ZIP code. There was some discrepancy in the number of ZIP codes for data sets from different sources. Census, ESRI, CARFAX, Polk, and ESRI sourcebook contain different numbers of ZIP codes. The ESRI ZIP codes were chosen as the primary data set. ZIP codes cases that appeared in other data sets were trimmed off. ESRI source book contained two separate tables of ZIP codes, one for residential zip codes, and one for non-residential ZIPs. The non-residential ZIPs were folded into the set containing residential ZIP codes using the latitude and longitude for the non residential ZIP, and the geographic boundaries for the residential ZIPs.

Hypothesis Table

Once all the data sources were finalized, an individual hypothesis was written for each variable. These hypotheses were based on existing literature where available, and, where the literature did not exist, the intuition of the experimenter. The hypothesis table below represents the variables that were found to create the most effective regression model.

Table 5: Hypothesis Table

VARIABLE	SIGN	HYPOTHESIS
% of Pop 15-64	+	Increases in the proportion of people of working age will result in increased VMT due to a greater proportion of the population commuting to work (Polzin, 2006)
% male	+	Increases in the percentage of the population that is male will result in increased VMT due to a greater proportion of the population commuting to work. Men make up 54.5% of the work force (U.S. DOL, 2007)
2009 Diversity Index	-	Increased racial diversity will result in decrease in VMT/C due to lower vehicle ownership (Contrino, 2009).
2009 % of people in families	-	Increases in the percentage of the population that is living in family units will result in decreases in VMT due to the effects of trip consolidation (Polzin, 2006)
2009 % of people group quarters	-	Increases in the percentage of the population that is living in group quarters will result in decreases in VMT/C . Colleges, universities, and military barracks, are conventionally believed to have lower VMT/C
2009 Average Family Size	-	Increases in the Average family size will result in decreases in VMT do to the effects of trip consolidation (Polzin, 2006)
2009 Population Density	-	Increases in population density will result in decreased VMT, as trip will be shorter, and walking more possible (Newman, 2006)

VARIABLE	SIGN	HYPOTHESIS
2009 Median Household Income	+	Increases in median household income will result in increases in VMT, as luxury trips are more available, and the cost of driving represents a smaller proportion of income (Polzin, 2006)
2009 % Housing Units Renter Occupied	-	Increases in the percentage of the housing units that are renter occupied will result in decreases in VMT, as renters have more available parking and because renting is rarer in typical sprawl situations.
2009 % Housing Units Vacant	+	Increases in the percentage of housing units that are vacant will result in increased VMT, as neighborhoods with vacant housing are seen as unsafe for walking (Nikerman, 2009)
2000-2009 Population: Annual Compound Growth Rate	+	Increases in population growth rate will result in increased VMT, as rapidly growing areas are more likely to experience sprawl like conditions (Nasser, 2001).
2000-2009 Median Household Income: Annual Compound Growth Rate	+	Increases in income growth rate will result in increased VMT, as they often indicate rural areas that are being taken over by suburban sprawl
% of 2009 Population 25+ with a Bachelor's Degree or Higher	-	Increases in the percentage of the population with a Bachelor's degree or higher will result in decreased VMT (Cervero, 2003)
Unemployment Rate	-	Increases in the unemployment rate will result in decreased VMT, as trips to work will decrease.
Total Jobs per Cap > 5 miles away	+	Increases in the number of jobs per capita between 5 and 30 miles from the ZIP code centroid will result in increased VMT, as available jobs are farther away and require longer commutes
Total Jobs per Cap <= 5 miles	-	Increases in the number of jobs per capita within 5 miles of the ZIP centroid will result in decreased VMT, as commuting distances are shortened.
Train Transit Available (by city)	-	The availability of train transit options within the City the ZIP falls under will result in lower VMT, as train trips replace car trips.
Bus Transit Available (by city)	-	The availability of bus transit options within the City the ZIP falls under will result in lower VMT, as bus trips replace car trips.

IV. Results

Results Summary

When the results of the full regression are sorted by absolute value of the standardized coefficients (from here on, 'Beta'), the relative predictive power of the variables included can be ranked. The Table 6 (Variables Sorted by Standardized Coefficient) shows these rankings.

Table 6: Variables Sorted by Standardized Coefficients

Rank	Variable	Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	2009 % of people in families	8245.40	6.86	0.157	0.000
2	2009 % Housing Units Renter Occupied	-6339.96	3.86	-0.155	0.000
3	Total Jobs per Cap > 5 <= 30 miles	0.60	0.03	0.124	0.000
4	2009 Average Family Size	-1970.68	104.57	-0.121	0.000
5	2009 % of people group quarters	-6257.03	6.22	-0.098	0.000
6	2009 Population Density	-0.10	0.01	-0.089	0.000
7	Bus Transit Available (by city)	-1428.62	138.33	-0.084	0.000
8	2009 Median Household Income	-0.02	0.00	-0.078	0.000
9	Unemployment Rate	-88.26	6.49	-0.078	0.000
10	% of Pop 15-64	82.58	0.08	0.070	0.000
11	% of 2009 Population 25+ with a Bachelor's Degree or Higher	-2669.24	4.03	-0.067	0.000
12	2000-2009 Median Household Income: Annual Compound Growth Rate	-339.55	33.58	-0.055	0.000
13	2000-2009 Population: Annual Compound Growth Rate	-113.83	11.20	-0.053	0.000
14	2009 % Housing Units Vacant	2225.16	2.79	0.051	0.000

15	% male	94.29	0.11	0.051	0.000
16	2009 Diversity Index	-10.80	1.49	-0.047	0.000
17	Train trips per Year (unlinked) (by city)	25.83	5.77	0.033	0.000
18	Bus trips per Year (unlinked) (by city)	2.26	6.14	0.004	0.713
19	Train Transit Available (by city)	-40.27	202.26	-0.001	0.842
20	Total Jobs per Cap <= 5 miles	-0.03	0.33	-0.001	0.933

The housing variables (% of people in families, % of people in group quarters, average family size , population density , % housing units renter occupied, and % housing units vacant) dominate the top of the list. Five out of six of these variables have the highest Beta of all variables. (Table 6) All six variables in the group have significance levels of .000 in the complete model as well.

The jobs grouping had inconsistent predictive power. Total jobs per capita greater than 5 but less than 30 miles had the third largest Beta value in the complete regression, and unemployment rate scored 9th, and was significant. (Table 6) However, total jobs per capita less than 5 miles was the least effective variable at predicting VMT, and had a significance value of .93, not even remotely significant. (Table 6)

Similarly, transit availability was widely spread, with bus transit availability producing the 7th largest Beta, but the other three being in the bottom four when ranked by absolute value of Beta. Additionally, bus trips per year and train transit availability were not significant in the complete regression. (Table 6)

Demographic variables tended to be middle of the pack, with income taking the highest spot of the group at 8th. Diversity and percent male were relatively ineffective at predicting VMT, scoring 15th and 16th respectively. (Table 6)

While nether of the growth rate variables were in the top half in terms of predictive power, the growth rate variables took spots 11 and 12, so were middle of the road. (Table 6)

Transit Availability & Trips

Table 7: Stepwise, Transit First

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.308a	0.095	0.095	5305.36
2	.486b	0.236	0.236	4874.00
3	.509c	0.259	0.258	4801.97
4	.527d	0.278	0.277	4739.43
5	.532e	0.283	0.282	4723.31

Coefficients					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	13380.10	33.00		0.000
	Train trips per Year (unlinked) (by city)	-28.13	5.48	-0.036	0.000
	Train Transit Available (by city)	-439.07	226.02	-0.015	0.052
	Bus trips per Year (unlinked) (by city)	-34.39	6.83	-0.055	0.000
	Bus Transit Available (by city)	-4098.10	148.58	-0.240	0.000

2 Housing added (Constant)		14132.77	522.26		0.000
	Train trips per Year (unlinked) (by city)	35.96	5.90	0.047	0.000
	Train Transit Available (by city)	-70.69	208.30	-0.002	0.734
	Bus trips per Year (unlinked) (by city)	2.79	6.31	0.004	0.659
	Bus Transit Available (by city)	-2024.84	141.26	-0.119	0.000
3 housing + Demographics added (Constant)		3689.34	899.92		0.000
	Train trips per Year (unlinked) (by city)	25.47	5.83	0.033	0.000
	Train Transit Available (by city)	13.53	205.50	0.000	0.948
	Bus trips per Year (unlinked) (by city)	2.97	6.23	0.005	0.634
	Bus Transit Available (by city)	-1611.67	140.44	-0.094	0.000
4 Housing + demographics + jobs added (Constant)		5461.16	896.55		0.000
	Train trips per Year (unlinked) (by city)	24.67	5.79	0.032	0.000
	Train Transit Available (by city)	-41.11	202.91	-0.001	0.839
	Bus trips per Year (unlinked) (by city)	1.53	6.15	0.002	0.804
	Bus Transit Available (by city)	-1446.25	138.76	-0.085	0.000
5 Housing + demographics + jobs + growth rate added (Constant)		6955.57	910.50		0.000
	Train trips per Year (unlinked) (by city)	25.83	5.77	0.033	0.000
	Train Transit Available (by city)	-40.27	202.26	-0.001	0.842
	Bus trips per Year (unlinked) (by city)	2.26	6.14	0.004	0.713
	Bus Transit Available (by city)	-1428.62	138.33	-0.084	0.000

In the stepwise regression starting with transit, all four transit variables have negative coefficients (Table 7: Stepwise, Transit First) when run together in isolation. Of them, however, bus availability is by far the strongest both in terms of standardized and unstandardized coefficients (from here on 'B'). The results of the isolated regression show that the availability of bus transit within a ZIP reduced VMT per capita for that ZIP by over 4,000 miles per year (Table 7). With the addition of housing variables, however, the signs on train trips per year and bus trips per year flip. Train transit availability and bus trips per year also become insignificant as predictors of VMT/C. With the inclusion of demographic variables, only bus transit availability is negative, and its effects are reduced to around a 1,600 reduction of VMT/C for ZIPs where buses are available. (Table 7) The train trips per year variable remains significant, and shows that for each additional train trip per person per year, VMT/C increases by 25 miles. (Table 7)

With all the variables included, the availability of train transit per year results in a 1,400 miles per year decrease in average VMT/C. The variable and train trips per person per year is essentially unaffected by the addition of job or growth rate variables and continues to show a 25 mile per year increase for each additional average train trip per person per year.

The overall effects of transit on VMT/C are complicated, and the results of this model only provide a glance at what is a multifaceted issue.

Use of the variables describing transit unlinked trips per year carries with it certain complexities. An increased number of trips per year on transit could have one of two implications:

- Transit trips are replacing vehicle trips in those areas
- The total number of trips needed for everyday functions is higher in those areas

If the first of these possibilities is true, then increases in transit ridership should result in decreases in VMT/C. Trips that would normally be made using a car, are instead moved to a bus or train. If the second possibility is true, then transit trips per year might belong with VMT/C as a dependent variable, and the research question would be 'what drives travel demand.' The results of correlation tests would seem to point to the first option being more probable, as there is only a -.12 Pearson's correlation between total unlinked trips and VMT/C (Appendix F). The variation in the variables, then is not is not highly correlated, enforcing the idea that they are not both similarly influenced by the total demand for travel.

However, the results of the regression would seem to favor the second explanation over the first. The number of bus trips per year was insignificant, so there is not much that can be gained from that. Train transit had a slightly positive effect on VMT/C: for every additional average trip per year, VMT/C also increased by approximately 26 miles per year. (Table 7) This result indicates that the second possibility is likely to be holding true. ZIP codes with increased demand for travel due to other factors experience both higher VMT and higher transit usage. Additionally, when run in isolation both bus and train trips have negative coefficients, but once housing characteristics are added into the regression the signs flip. (Table 7) This can be interpreted as indicating that while the overall effects of transit usage result in a drop in VMT, this decrease is correctly attributable to the housing characteristics of areas with transit, not the ridership rates on the transit itself. The housing characteristics, then, drive increased trip usage, some of which is in the form of increased VMT/C, and some of which is increased transit ridership.

The result of the inclusion of transit availability as well as trips per year adds another element to the analysis. Train transit availability in isolation resulted in a small decrease in VMT/C. However, once housing was added into the regression the variable for train transit availability became insignificant.(Table 7) Any reduction in VMT/C from the isolated regression associated with train transit availability, then should instead be attributed to the housing characteristics that come with train availability. Bus availability, however, does indicate that having access to bus based transit options reduces VMT/C, even when all the other variables tested in this project are included. (Table 7) Bus availability remains a significant factor, and as the coefficient has a negative sign, increases in bus availability equate to lower VMT/C.

The transit data has some unique issues. One of the advantages this study has over previous studies is the small size of the geographic unit of analysis. Transit ridership and availability data are only available at the city level, and had to be translated down to the ZIP code level. This process eliminated this advantage of the study and means that all the findings on transportation should be viewed in light of this information.

Jobs

Table 8: Stepwise, Jobs First

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.220a	0.048	0.048	5439.53
2	.425b	0.181	0.180	5047.72
3	.522c	0.273	0.272	4755.66
4	.527d	0.278	0.277	4739.43
5	.532e	0.283	0.282	4723.31

Coefficients				
Model	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	

1	(Constant)	14662.24	70.78		0.000
	Unemployment Rate	-204.50	6.46	-0.180	0.000
	Total Jobs per Cap > 5 <= 30 miles	0.62	0.03	0.128	0.000
	Total Jobs per Cap <= 5 miles	-5.89	0.37	-0.106	0.000
2 Demographics added (Constant)		19777.66	576.23		0.000
	Unemployment Rate	-173.55	6.64	-0.153	0.000
	Total Jobs per Cap > 5 <= 30 miles	0.64	0.03	0.131	0.000
	Total Jobs per Cap <= 5 miles	-2.88	0.35	-0.052	0.000
3 Demographics + Housing added (Constant)		4017.72	892.29		0.000
	Unemployment Rate	-91.50	6.52	-0.081	0.000
	Total Jobs per Cap > 5 <= 30 miles	0.63	0.03	0.130	0.000
	Total Jobs per Cap <= 5 miles	-0.30	0.33	-0.005	0.373
4 Demographics + Housing+ transit added (Constant)		5461.16	896.55		0.000
	Unemployment Rate	-88.15	6.51	-0.078	0.000
	Total Jobs per Cap > 5 <= 30 miles	0.62	0.03	0.127	0.000
	Total Jobs per Cap <= 5 miles	-0.28	0.34	-0.005	0.408
5 Demographics + housing + transit + growth rate added (Constant)		6955.57	910.50		0.000
	Unemployment Rate	-88.26	6.49	-0.078	0.000
	Total Jobs per Cap > 5 <= 30 miles	0.60	0.03	0.124	0.000
	Total Jobs per Cap <= 5 miles	-0.03	0.33	-0.001	0.933

The variables in the job category, when regressed together produced an R square of .048 (Table 8: Stepped Regression, Jobs First) . In isolation the variable for total jobs per capita between 5 and 30 miles had a small (B = .62) but positive unstandardized coefficient, which is in line with that hypothesis. (Table 8) More jobs between 5 and 30 miles away per resident of the ZIP code is expected to result in higher driving needs, as commutes are longer. The Beta of .128 indicates that this effect is relatively strong. (Table 8) The variable jobs within 5 miles per resident of the ZIP produced a B of -4.9, and a Beta of -.12. This also is in line with the hypothesis, as increased density of jobs close to places of residence should decrease the need to drive to work, and the distance driven. (Table 8)

Controlling for the effects of basic demographic features, the effects of jobs remain relatively constant, however the Beta of jobs less than 5 miles decreases somewhat. This steadiness indicates that the effects of job density and location are not closely linked to demographics. However, with the addition of housing stock variables, total jobs per capita less than 5 miles loses significance as a predictor variable of VMT/C. (Table 8) The effects of unemployment rate are reduced by the addition of demographics and housing stock variables, but remain significant.

Addition of controls for transit availability and growth rate have no important impact on the effects of job location. Total jobs less than 5 miles remains insignificant, while total jobs between 5 and 30 miles and unemployment rate stay essentially unchanged. (Table 8) The results of the complete regression show that for each increase of one percent in the unemployment rate, VMT/C decreases by about 88 miles per year, and that for each additional job within 30 miles but farther than 5 miles, per resident of the ZIP code, VMT/C for the zip increases by .6 miles per year. (Table 8)

Multiple, complicated methods of measuring the effects of job location and type were explored (Appendix H) However, in the end the most simplified variable options were chosen. Even using the simplified variables, the effects of jobs on VMT are not straight forward. The effects of total jobs between 5 and 30 miles from the ZIP code are easily readable from the models. Greater concentrations of jobs in this area results in increased VMT/C. Even with the addition of all the other variables, this effect is quite steady; across all models, for every additional job per worker in the ring, VMT/C increases by .6 miles per person per year. (Table 8) This finding is in line with the hypothesis, as the more jobs that exist away from the center of the ZIP code, the more likely it is that a commuter will be commuting that distance to work. When one looks at the number of jobs per worker within 5 miles of the ZIP centroid, the interpretation becomes more complicated. When run with just the other job variables, jobs per worker within 5 miles has a Beta of -.106, and a B of -5.89. This is consistent with expectations, as more jobs closer is both an indication of mixed use - a type of development believed to reduce VMT - and more directly means that commuters need not commute as far to work. (Table 8) The pattern stays the same with the addition of demographic variables into the regression, though the variable loses some predictive power. However, with the addition of housing characteristic, the effect becomes insignificant. (Table 8) Identifying the variable or variables within the housing characteristics group that exert such influence on the jobs within 5 miles variable is difficult, as the normal indicators, such as high correlation values, are not there. (Appendix F)

It is logical that the jobs within 5 miles variable would be more affected by the entrance of other variables than the job from 5 - 30 miles variable. The jobs within 5 miles variable include the jobs within the zip code. Especially in large, rural ZIPs, the 5 mile ring may not cover many, or any other ZIP codes. That the jobs within a ZIP code are affected by other factors describing the ZIP code is not surprising. Conversely, that the coefficients for the ring for jobs 5-30 miles away from the center of the ZIP is changed very little by the inclusion of other variables is to be expected.

The other variable explored in the jobs category was unemployment rate. As hypothesized, the unemployment rate exerted a suppressive effect on VMT. These effects held through the inclusion of more variables. The addition of demographic variables, including income, to the models had a lower predictive power reduction than expected. Income and unemployment rate were not even highly correlated - with a Pearson's value of -.237. (Appendix F) This means that the effects of unemployment on VMT are separate from these other factors. The decrease in VMT that comes with increased unemployment rate, then, is due to factors such as decreased commuting time, not just decreased income.

Housing

Table 9: Stepwise, Housing First

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.308a	0.095	0.095	5305.36
2	.486b	0.236	0.236	4874.00
3	.509c	0.259	0.258	4801.97
4	.527d	0.278	0.277	4739.43
5	.532e	0.283	0.282	4723.31

Coefficients				
Model	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	

1	(Constant)	12960.69	512.81		0.000
	2009 % of people in families	117.83	6.22	0.224	0.000
	2009 % of people group quarters	10.87	5.87	0.017	0.064
	2009 Average Family Size	-2783.55	100.20	-0.172	0.000
	2009 Population Density	-0.12	0.01	-0.100	0.000
	2009 % Housing Units Renter Occupied	-77.40	3.48	-0.190	0.000
	2009 % Housing Units Vacant	36.82	2.56	0.084	0.000
2 Demographics added	(Constant)	2028.05	895.13		0.023
	2009 % of people in families	117.80	6.66	0.224	0.000
	2009 % of people group quarters	-27.15	6.04	-0.042	0.000
	2009 Average Family Size	-2472.79	104.05	-0.152	0.000
	2009 Population Density	-0.08	0.01	-0.071	0.000
	2009 % Housing Units Renter Occupied	-72.17	3.82	-0.177	0.000
	2009 % Housing Units Vacant	21.17	2.82	0.048	0.000
3 Demographics + Jobs added	(Constant)	4017.72	892.29		0.000
	2009 % of people in families	111.42	6.62	0.212	0.000
	2009 % of people group quarters	-38.02	5.97	-0.059	0.000
	2009 Average Family Size	-2197.73	104.37	-0.135	0.000
	2009 Population Density	-0.08	0.01	-0.067	0.000
	2009 % Housing Units Renter Occupied	-69.48	3.79	-0.170	0.000
	2009 % Housing Units Vacant	24.22	2.80	0.055	0.000
4 Demographics + Jobs + Transit added	(Constant)	5461.16	896.55		0.000
	2009 % of people in families	97.95	6.74	0.186	0.000
	2009 % of people group quarters	-48.12	6.09	-0.075	0.000
	2009 Average Family Size	-2120.98	104.34	-0.131	0.000
	2009 Population Density	-0.10	0.01	-0.086	0.000

	2009 % Housing Units Renter Occupied	-61.68	3.83	-0.151	0.000
	2009 % Housing Units Vacant	22.34	2.80	0.051	0.000
5 Demographics + Jobs + Transit + Growth rate added (Constant)		6955.57	910.50		0.000
	2009 % of people in families	82.45	6.86	0.157	0.000
	2009 % of people group quarters	-62.57	6.22	-0.098	0.000
	2009 Average Family Size	-1970.68	104.57	-0.121	0.000
	2009 Population Density	-0.10	0.01	-0.089	0.000
	2009 % Housing Units Renter Occupied	-63.40	3.86	-0.155	0.000
	2009 % Housing Units Vacant	22.25	2.79	0.051	0.000

The variables that make up the housing stock and population distribution category are some of the most influential variables in the model (Table 5). Five of the top six variables, sorted by absolute value of Beta, fall under this grouping. The R squared for the housing variables in isolation is .226 (Table 9: Stepped Regression, Housing First). As a matter of comparison, the R squared for the complete model is .282. Eighty percent, then, of the predictive power of the model is found in the housing stock and population distribution variables.

The living arrangements of people across the United States can be divided into three groupings, those living in family units, those living in non -family households, and those living in group quarters. For the sake of this analysis, the percent of the population living in non-family households was removed to reduce multicollinearity. (Appendix G) The percent of the population living in families consistently shows up as the most powerful predictor of VMT/C used in the model. (Table 5) When the housing stock and population distribution variables are run in isolation, the model output says that for every increase of one percent in the proportion of people living in families, VMT/C increases 118 miles per person per year. (Table 9) Interestingly, the percent of people living in group quarters, when run with only the other housing variables is insignificant. However, once more variables are added to the model, the variable becomes significant. (Table 9)

Another housing stock descriptive variable is the percentage of residences that are owner occupied, renter occupied, or vacant. For this analysis the variable percentage of the population owner occupied was removed to reduce multicollinearity. Housing units renter occupied had a B of -77, or a change of 77 miles for each change of one percent. (Table 9) It is also negative, meaning that higher proportions of renter occupied housing decreases VMT/C. The percentage of housing units that are vacant has a positive B coefficient of 37, so with just the other housing and population distribution variables accounted for, the model shows that higher vacancy rate result in increased VMT. (Table 9)

The other two variables in the housing and population distribution group are average family size and population density. Average family size shows a 27 VMT/C per year reduction for the increase of the average family size by one person. Population density produces a reduction of .12 VMT/C per year for each increase of one unit, or one person per square mile. (Table 9)

The addition of demographic variables to the model has one pronounced effect on the housing variables. The percentage of people in group quarters, which previously was insignificant and had a positive sign, becomes significant, and caries a B coefficient value of -27. (Table 9) All the other variables remain essentially unchanged by the addition of demographic data. The addition of job variables again has minor weakening effects on most of the housing variables, but strengthens the pull of the percent of people in group quarters. This trend continues with the addition of transit and growth rate variables. (Table 9)

The variables in the housing group define the type, use, and location of the living conditions found within a ZIP. The explanation of the variables used in the regression can be found in Table 10.

Table 10: Housing Group Variables

GROUP	VARIABLES	USE
Living situation	• % people in families • % people in group quarters • % people in non-family households (excluded)	These variables describe the distribution of people within housing structures. Group quarters populations are often colleges, universities, and barracks.
Housing type	• % units renter occupied • % units vacant • % Units owner occupier (excluded)	These variables describe the physical housing units. A higher percentage of renter occupied units is an indication of more apartments and multi-family homes. The percentage of units vacant is an indicator of economic conditions, but also the walkability of the streets
Within housing unit density	• Average Family Size	This variable further describes the distribution of people within housing units. Additionally larger groups living together are hypothesized to reduce VMT through trip sharing
Housing unit physical arrangement	• Population density	This variable describes the physical arrangement of housing units. It is thought to be one of the driving factors of VMT usage, because of decreased distances between services and the availability of transit alternatives

These factors have been shown to be the most important in predicting VMT/C. The percentage of people living in families versus group quarters (versus the third segment of the whole: non family households) shows that traditional family units cause more driving than atypical living situations, such as group quarters or non-family groups. (Table 9) It was hypothesized that increased proportions of families would decrease VMT/C because children share rides with their parents. This hypothesis was misguided and poorly thought out. The effects of age on VMT/C are already controlled by other variables. Furthermore, family units are associated with sprawling developments, as people move out of the cities to raise children in more pastoral conditions. The stereotypical figure of the mini-van bound soccer mom should have also been an indication that families are the quintessential high VMT/C population.

The effects of the number of people in group quarters is magnified as more variables are added to the regression. With every addition of a group of variables, both the B and the Beta of the variable become increasingly large. This means that the effects of people in group quarters are uniquely attributable to the variable. When only the housing variables are included, an increase in percentage of the population in group quarters actually leads to an increase in VMT, however once demographics are added to the regression these effects switch directions. (Table 9) This means that those who live in group quarters possess characteristics that would otherwise be associated with high VMT/C. Because living in group quarters suppresses this driving, the more variables in the regression, the more profound the effects are. Colleges and universities are major locations of group quarters populations. The age breakdown of colleges and universities is almost exclusively in the 15 - 64 range. However the unique situation of college campuses means that even though there may be a lot of working age people, not a lot of people work, and those who do are less likely to drive to work.

The renter occupied versus owner occupied versus vacant variables are relatively unaffected by the other variables in the model. This is unexpected, as previous research would lead to the belief that these variables would be heavily influenced by demographic factors such as income and age. Indeed the addition of the demographic variables does affect these variables more than the addition of any other group, but the effects are still minor.

There exists a large body of other research showing that increased density results in lower VMT/C. It was not surprising, then to find that result constantly in these regressions. There are a plethora of reasons that higher densities enable lower VMT/C. For this study the most important part of the inclusion of density into the models was to be able to account for all these effects proven to be linked to density, and explore if other factors had unique effects, or if they were merely tied to VMT/C through the connection to density. As many variables had significant effects even with the inclusion of density, it can be said that the forces driving VMT/C are too complicated and diverse to be easily summed up with one, or even a few variables.

Other Effects of Note

<u>Income</u>

Demographic data in this analysis were used more as a control than as primary exploratory variables. Thus as demographics are difficult to control, it was less important to this study to understand their effects in isolation, than it was to understand how the demographics of a ZIP code influence the factors that can more easily be changed.

The effects of demographic variables, however, exist and play a large part in the success of the model. Income has the 8th largest Beta of any variable in the complete regression.(Table 5) For every additional dollar of average income, VMT/C decreased by .02 miles per year. (Table 5) This seems minor when examined at that scale, but one standard deviation of income is $21,000, so an increase of one standard deviation equals a 440 miles per year decrease in VMT/C. (Appendix E)

The finding that higher income levels results in lower VMT/C was one of the most surprising of this study. The previous literature is split on the issue of whether income is a factor, however none of the studies reviewed suggested that higher income could suppress VMT/C.

Age

The breakdown of age in a ZIP code is also a significant indication of VMT/C patterns. If a larger proportion of the population is in the 15 - 65 age group that is likely to be actively driving, VMT/C increases. (Table 5)

Growth Rate

In the final regression, the fact that both growth variables had negative coefficients has interesting implications. It was hypothesized that population growth rate in particular would have a positive coefficient, as the fastest growing areas are often sprawl. Examining the correlations between growth rate and the other variables does not shed much new light on the issue, as both growth variables are hardly correlated with the other variables at all, and what correlation there is (.1 with average family size, and -.1 with % people in group quarters) would seem to point in the opposite direction. One possible explanation is that there is a problem with the methodology. The data for a rapidly growing ZIP code might be thrown off by the rapid change. All efforts were made to ensure that every variable used represented a snapshot of the country at the same time, however there is some window over which the data was collected. If the Polk data on the number of vehicles was a little older than the population data, rapidly growing areas would appear to have lower VMT/C and the population would be artificially large in comparison to the number of vehicles.

Diversity

Diversity, while a complicated issue, also has enlightening effects on VMT/C. The diversity index variable used in the final regression has a - comparatively weak, but still completely significant - suppressive effect on VMT/C. There are two explanations for why this may be the case.

- Diversity and low VMT/C are both associated with urban areas. While the model controls for density, density is not a perfect predictor of an area being urban, so it is to be expected that if density and urbanism are closely related, and if urban environments do have lower VMT/C per capita even beyond the effect of density, than some of these effects will be captured by the diversity variable.

- Racial differences, and, more specifically the cultural norms of different ethnic groups, affect driving habits. Results from a regression run with racial groups by percentages shows that there may be some driving habit difference, though these are obscured when all the variables in the model are accounted for. While almost none of these results were significant when included in the whole model, there are trends towards some groups driving more than others.

The diversity index variable would seem to support the first conclusion, while the independent racial percentages supports the second. (Appendix I) As is so often the case with this type of thing, however, the best explanation likely borrows truths from both.

Education Levels

Only the variable representing the percentage of people with a bachelors degree or higher was included in the complete regression. This variable showed that those with the highest level of education - about 20% of the U.S. population has a bachelors degree or higher - drive less. The effects of education level are undoubtably far more complicated then this analysis allows.

VI. Implications

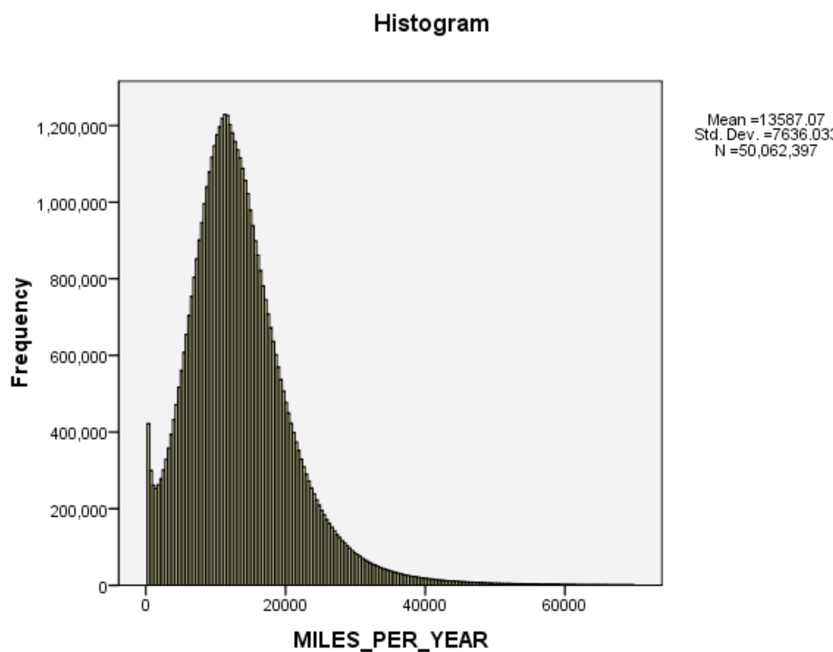
Effective Strategies for VMT Reduction

Uses of the Findings

Understanding how jobs, housing, transit, demographics, and growth rate variables influence the driving needs of a ZIP code have implications for city planners. Better models to predict VMT based on other patterns of change could also be useful to other federal agencies, local governments, and groups looking to predict how American driving habits will evolve.

As importantly, being able to predict the VMT/C of an area based on easily obtainable data can assist local governments considering new development projects. More complete information as to how much a given proposal will add to the VMT/C and therefore road infrastructure costs can give a better idea of what the true costs of a proposal are, and may stop sprawling projects, which might appear cheeper but may cost more in the long run due to increased infrastructure demands. Many advocacy groups espouse the general concept that sprawling development costs more to maintain, partially due to in-

Figure 11: Distribution of VMT



creased road use (New Urbanism, 2010). The ability to put an exact figure on this, as this paper does, can add validity to this argument.

While low density is not the only feature of sprawl, it is an important aspect, and this study demonstrates that for every increase of population density by one person per square mile, VMT/C decreases by 0.1 miles per year. When put in perspective, this fact is slightly more compelling. The mean population density in the United States is 1,266 people per square mile. An area with double the density would then have 127 fewer miles traveled per capita per year. The average VMT/C for the United States is 12,875 miles per year, so a doubling of density equates to a 1% drop in VMT/C. While still not a huge number, the standard deviation of density is large -- 4,763 people/sqr. mile -- so a doubling in density is a relatively small increase.

Especially with the possibility of tighter controls on carbon emissions moving forward, accurate understanding of how policies can affect VMT can help areas meet carbon goals. It is even possible, though not likely in the near future, that implementation of policies proven to reduce VMT can produce carbon credits for a future carbon market.

Most Effective Strategies

Overall, based on the findings of this study, it would appear that policies aimed at affecting the housing structure of an area will be the most effective at reducing VMT. Encouraging larger groups of people to live together in denser areas is a more effective way of controlling VMT than the introduction of additional transit options. These finding imply that funding should be shifted to incentivize this type of living situation rather then put towards transit, in particular train based transit.

Some Variables are Difficult to Control

Not all of the variables found to have strong effects on VMT/C are easily controlled. The social structure of families is not something that should be changed to lower VMT/C.

As increased family units are responsible for a large proportion of VMT growth, identifying exactly what factors make a family require more driving than their non-family peers is an effective step in tackling the issue. It does not seem an advisable course of action for society to discourage the formation of family units, but enacting policies that shift the factors unique to families more into line with those of less driving intensive groups such as residents of group quarters would have a large impact on VMT reduction.

Focus on Bus Transit

While not as important as other factors, the availability of bus based transit did lower VMT/C. Areas looking to invest in transit for the purpose of decreasing driving should focus efforts and money on systems that favor busses over trains.

Avoid Building Away From Job Hubs

The more jobs there are, relative to the population of the ZIP code, between five and thirty miles away, the higher VMT/C is. The results indicate that such development patterns as placing new developments on the fringe of urban cores will raise the VMT/C of those developments.

Density Helps

While not an overwhelming factor as some previous studies would seem to suggest, denser areas to have lower VMT/C. Density is easy to control, so should be one of the first things a planner looks at when designing an area to have lower VMT/C. (Table 5)

Higher Income Helps

While income is not a variable that can be consciously manipulated to lower or raise VMT/C, understanding the effects of income on the VMT/C is very useful in accurately predicting how other variables behave.

The finding also calls into question some of the basic principles that lead researchers to believe income would have a positive effect on VMT/C. If higher income makes the costs of driving more palatable and increases the number of luxury activities that involve trips, as has been hypothesized, then why does having a higher income suppress VMT/C? One possible explanation is that wealth enables one to live closer to jobs and other activities. Those in the lower income brackets tend to be forced to the outskirts of cities where land is cheeper. Affordable housing in central locations, then, may be an important part of development design intended to lower VMT/C.

Take Lessons from College Campuses, Barracks

While it is unreasonable to expect everyone to move into a dorm, there seems to be a lot to learn about the channels through which living in group quarters suppresses VMT/C. It is beyond the scope of this study to suggest what the channels may be, but further study into what exactly it is that makes living in group quarters so much less driving intensive is warranted. There are likely lessons can be derived from group quarters that may be applied to more traditional housing types.

Favor Rentable Construction

While this study does not explain how it is that the greater the proportion of the population that rents their homes, the lower VMT/C is, the fact that this relationship exists says that more focus should be put on building construction intended for rental.

Implications for Existing Urban Design Policy

Smart Growth

The Smart Growth Network is an organization that works on the national, regional, and local levels to support coalitions for intelligent and sustainable growth. Smart Growth has 10 guiding principles for how the organization thinks development should occur:

- Create a range of housing opportunities and choices
- Create walkable neighborhoods
- Encourage community and stakeholder collaboration
- Foster distinctive, attractive communities with a strong sense of place
- Make development decisions predictable, fair, and cost effective
- Mix land uses
- Preserve opened space
- Provide a variety of transportation choices

• Strengthen and direct development towards existing communities

• Take advantage of compact building design (Smart Growth, 2010)

The results of this study have direct implications for three of these principles. The findings regarding renting vs. owning, and family vs. group quarters support the principle that a range of housing opportunities should be provided. Atypical housing situations were shown to drive lower VMT/C, so planned development in which the intent is to encourage less driving, should encourage theses settlement patterns.

The results of the jobs variables in the regression support -- to some degree -- the idea of mixed land use. One of the reasons that Smart Growth supports mixed land use is that it believes having jobs closer to housing shortens commutes. The results of this project support that hypothesis, as areas where the job concentration from 5-30 miles away was higher, experienced higher VMT/C.

Finally, the principle that a variety of transportation choices is beneficial to an area is somewhat supported by the findings of this study. Bus transit does reduce VMT/C where available. However, access to train transit did not have a significant effect.

The principles of Smart Growth are not aimed at reducing VMT/C so much as they are at improving general quality of life. Reduction of VMT/C is just one of many pathways through which they propose that quality of life can be improved by neighborhood design.

<u>New Urbanism</u>

Like Smart Growth, New Urbanism is a framework that covers more than just reducing car use. New Urbanism, however, has a particular focus on density and train based transit, both factors for which this study has implications.

New Urbanism pushes high density because the organization believes that higher density areas allow for increased walking, decreased driving, beautiful public spaces, and a more convenient lifestyle (New Urbanism, 2010). This study supports one aspect of this claim. Higher population densities were found to suppress VMT/C.

The focus on train based transit, however, is not supported by the results of this study.

VII. Conclusion

Finding Limits

The primary results of this project are based on regression analysis. This allow for confidence in correlation, but not the causality of the results. Previous studies informed the hypothesis for the project, and allow for informed speculation into the direction of the effects. The use of direct correlation analysis in conjunction with regression findings further inform the conclusions drawn.

The aim of this project is not to determine *why* the variables used affect VMT/C, but instead to tell how much and in what direction the variables exert influence.

Further Study

This analysis is a foundation, and a sample of what can be done with the data that exist. There is potential for much more complicated analysis of many of these factors. The use of combined and overview variables in this study added to the interpretability of the results, but may have obscured some of the underlying trends.

The data on transit is less robust than the rest of the data used. Sufficient data on transit availability and ridership are not available on the ZIP code level, so the values for this study do not have quite the predictive power they might otherwise have.

This study raises more questions than it answers. The purpose of the study is to create general rules regarding how jobs, housing, transit, demographic makeup, and growth rate influence the VMT/C of residents living within a given ZIP code. It was not within the scope of this study to answers questions such as: “Why does an increase in the precent of people who rent their houses result in lower VMT/C?”

The intent of this study is to establish the grounding of a new type of VMT study in which the aggregate VMT/C data are no longer a limiting factor. The removal of this constraint opens up many new pathways and methodologies for researchers. This project started the exploration down what might be considered the most obvious of such new directions; big picture studies that seek to explain the general trends of what factors influence VMT/C.

The company CARFAX retains ownership of the VMT data used for this experiment. They will, however, be willing to work directly with other researchers who are interested in utilizing their database to conduct research into VMT related issues.

There is rich opportunity for further research in this field, particularly with the use of CARFAX’s data. Some of the findings of this project are particularly ripe for future research.

Family Structure

The proportion of the population living in family units and the average size of families had large impacts on VMT/C. While useful to know, what is more relevant to the goal of reducing VMT/C is to understand why living in family units vs. group quarters increased VMT/C and why living in larger family units decreases it.

The more people living under one roof, the easier it is to combine trips, in theory. The principle of combining trips may also be applicable to children. A possible explanation for the effects observed in the study is that, for example, two children do not require twice as much driving as a single child. That is to say, there may be some vehicular economy of scale with larger families.

One possible methodology is to narrow the focus down to a few different areas and control for variables such as the number of vehicles per person, and the percentage of the population with drivers licenses. A survey based methodology that asked questions relating to why trips were taken, and the number of people in the car for varying types of trips could also help clarify the issue.

<u>Renter Occupied vs. Owner Occupied Dwellings</u>

This study found that having a higher percentage of homes renter occupied resulted in lower VMT/C. There are a number of possible explanations for this, but significant further study is required to move beyond the point of speculation.

It can be postulated that the effects of rentership on VMT/C are related to factors unique to renting. One possibility is that renters are more likely to be younger, more transient, and less firmly established. Another possible explanation is that renters have lower car ownership rates. This study does not have data to ether confirm or reject ether of these possibilities.

A way to help determine if this drop in VMT/C is caused by increased mobility of renters -- that they can chose residences closer to their places of work or frequent non-work destinations -- is to study the move frequencies of renters and their average trip lengths to work and non-work destinations. Higher movement frequencies and shorter work and/or non-work trip lengths would indicate that renters did indeed benefit from the ability to shift home addresses.

Other factors possibly involved include parking availability, car ownership rates, and neighborhood walkability. A study with a small geographic area where such factors can be easily examined could help answer these hypothesis's and get to the question of why rates of rental vs. ownership so greatly affect VMT/C.

Jobs

The effects of job location and type were only cursorily examined in this study. Future studies on the subject would benefit the state of knowledge by breaking out different types of jobs. It might be expected, for example, that a greater concentration of retail jobs would have a different affect on VMT/C of an area than a high concentration of office jobs. A methodology that is able to isolate the different types of jobs and also include their respective distances from the population in question would be better able to explain how the location and type of jobs affects VMT/C.

Mixed use is one of the keywords of development ideologies such as Smart Growth. A metric that measures the mix of housing and the variety of different types of jobs in an area might be able to determine what affect, if any, building in a way classified as ‘mixed use’ has on VMT/C.

Transit

There already exists a huge body of literature on transit and how it affects VMT/C. The most important implication of this study for the subject is simply that a database of detailed VMT/C records exists. Transit studies have previously been limited to areas where VMT data exists. The availability of VMT data for all areas across the country means that new examples of transit can be studied.

While the results of this analysis of transit are not clear, a few general trends exist:

- The ridership of transit, specifically train-based transit, likely behaves along with VMT/C as a response to transportation need pressures exerted by other factors.
- The presence of bus-based transit does serve as a reducer of VMT/C; access to bus based public transportation is likely absorbing travel that would otherwise be car trips.

There are limits to how far these conclusions can be drawn, and a lot of potential for future research. The model does not control for individual selection preferences in terms of living location, so the possibility exists that people who are inclined to drive less move to areas with bus access. The model also does not include vehicle ownership rates as a factor, because its effects are massive and dilute the predictive power of all other variables in the model.

In order better understand why access to transit effects VMT/C as this study shows it does, future research should focus on further isolating the effects of transit. The use of survey or other more qualitative data to explore why people ride buses and trains, and when and where they go can also help elucidate some of the questions around these results.

The addition of a new rail stop, in, for example Pelham NY, can be studied by looking at the average VMT/C before and after. Alternatively, instead of using the temporal before-after approach, studies could identify areas across the country that are similar but for the availability of transit. As VMT exists for all areas, studies are no longer limited to examining only certain select areas.

Education

Education in this study was accounted for using only a single variable of the percentage of the population that had completed college. The greater the proportion of the population with a college education or greater, he lower VMT/C. The causes of this could include:

- Ties to grad schools and research campuses. Nearly all residents in such ZIPs have a high degree of education, and the campus structure lowers VMT/C.

- There is some factor of education level itself that encourages lower VMT/C.

Preliminary regressions were run that explored alternative education variables. Interestingly, the higher the percentage of the population with at least a high school education, the higher VMT/C is. Somewhere between a high school and bachelors degree, then is the education level where VMT/C is highest. The relationship between education level and VMT/C is not linear, and is likely to be influenced by different factors at different levels. This conclusion is strengthened by the results when education level is entered into the model as a continuous variable, as it does not provide much predictive power.

It is expected that the actual effects of education are much more complicated. A more detailed education level variable was originally explored for this study, however the simplified one was chosen to minimize collinearity. The data on education levels is widely available from both the census and commercial sources.

Future studies on the issue would need a methodology that allows for a more detailed breakdown of education levels, and is robust to collinearity.

Vacancy Rates

High vacancy rates suppress VMT/C. (Table 9) This effect holds true even controlling for the effects of income. The vacant houses themselves may be exerting pressures to raise VMT/C. Vacant houses may make streets less attractive for pedestrian traffic. The effect could also be going the other direction. It may be that the areas with the highest vacancy rates are those where VMT/C is high for other reasons.

One possibility is that the cost of transportation in some areas is driving people out, and into other areas where it is easier and cheaper to get around. High vacancy rates, then may be a result of high VMT/C. More research is needed to determine the direction and causes of these results.

In order to better understand the effects of vacancy rates on VMT/C, studies assessing the walkability of areas with varying levels of vacancy should be conducted to determine if it is decreased walkability that makes areas with higher vacancy rates also have higher VMT/C. Cost of transportation studies can be conducted using more qualitative methods to determine the direction of this effect, and explore the idea that areas with high VMT/C have high vacancy rates because people are attempting to leave such areas due to the cost or undesirability of driving.

Works Cited

Burchell, R. W. and S. Mukherji (2003). "Conventional Development Versus Managed Growth: The Costs of Sprawl." Am J Public Health **93**(9): 1534-1540.

Burchell, R. W., A. United States. Federal Transit, et al. (2001). Costs of sprawl--2000. Washington, D.C., National Academy Press.

Cervero, R. and K. Kockelman (1997). "Travel demand and the 3Ds: Density, diversity, and design." Transportation Research Part D: Transport and Environment **2**(3): 199-219.

Contrino, H. a. M., Nancy (2009). "Demographics Matter Travel Demand, Options, and Characteristics Among Minority Populations " ublic Works Management Policy **13**(361).

DeLuchi, M. A. (1991). "Emissions of Greenhouse Gases from the Use of Transportation Fuels and Electricity." Center for Transportation Research **1**.

EIA. U.S. Energy Information Administration: Independent Statistics and Analysis. 2004

EIA, U. S. (2009). "United States Energy Profile." Retrieved 9/25/2009, 2009, from tonto.eia.doe.gov/country/country_energy_data.cfm?fps=US.

ESRI (2009). ESRI Demographic Update Methodology: 2009/2014. Redlands ESRI. **White paper**.

Ewing, R. P., Rolf; Chen, Don (1994). Measuring Sprawl and its Impacts, Smart Growth America.

Gee, G. C. and D. T. D. T. Takeuchi (2004). "Traffic stress, vehicular burden and well-being: A multilevel analysis." Social Science & Medicine **59**(2): 405-414.

Handy, S. (1996). "Methodologies for exploring the link between urban form and travel behavior." Transportation Research Part D: Transport and Environment **1**(2): 151-165.

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

Hillman, M., Ed. (1997). Cycling as the most realistic substitute for car use in urban areas: burying the conventional myth about public transport. The Greening of Urban Transport: Planning for Walking and Cycling in Western Cities. London, John Wiley and Sons.

McCubbin, D. R. a. D., M. A. (1996). The Contribution of Motor Vehicles and Other Sources to Ambient Air Pollution. Davis, Institute of Transportation Studies, University of California.

McCulloch, R. F., Philip; Darmawan, Jessica (2009). Getting on Track: Record Transit Ridership Increases Energy Independence, Enviroment America.

Mokhtarian, P. L. and I. Salomon (2001). "How derived is the demand for travel? Some conceptual and measurement considerations." Transportation Research Part A: Policy and Practice **35**(8): 695-719.

Nasser, H. E. O., Paul (2001). A comprehensive look at sprawl in America. USA Today.

New Urbanism. “Creating livable and sustainable communities” http://www.newurbanism.org/ 2010

Newman, P., & Kenworthy, Jeffrey (2006). "Urban Design to Reduce Automobile Dependence." Opolis.

Newman, P. a. J. R. K. (1999). Sustainability and cities: overcoming automobile dependence Washington DC, Island Press.

Newman, P. W. G. a. J. R. K. (1989). Cities and Automobile Dependence: An International Sourcebook. Aldershot, UK.

Polzin, S. E. (2006). The Case for Moderate Growth in Vehicle Miles of Travel: A Critical Juncture in U.S. Travel Behavior Trends. Tampa, Center for Urban Transportation Research, University of South Florida.

Pushkarev, B. a. J. Z. (1977). "Public Transportation and Land use Policy." Indiana Univeristy Press.

Smart Growth Network. “Smart Growth online” www.smartgrowth.org. 2010

Spash, C. L. (1997). "Assessing the Economic Benefits to Agriculture from Air Pollution Control " Journal of Economic Surveys **11**(47 -70).

Steg, L. "Car use: lust and must. Instrumental, symbolic and affective motives for car use." Transportation Research Part A: Policy and Practice **39**(2-3): 147-162.

Stradling, S. G., M. L. Meadows, et al. (2000). "Helping drivers out of their cars Integrating transport policy and social psychology for sustainable change." Transport Policy **7**(3): 207-215.

Transportation, U. S. D. o. (2003). "Annual Automobile VMT per Capita and Number of Automobiles per Capita 1997." Retrieved December 20, 2009, from http://www.nps.gov/history/nR/publications/bulletins/suburbs/part2.htm#figure2.

U.S. Department of Labor, Bureau of Labor Statistics, Employment and Earnings, 2008 Annual Averages and the Monthly Labor Review, November 2007.

U.S. Department of the Interior, N. P. S. "National Register Bulletin: Historic Residential Suburbs." Retrieved December 19, 2009, from http://www.nps.gov/history/nR/publications/bulletins/suburbs/part2.htm#figure2.

Appendices

List of Appendices:

To be useful for this project the data need to be in the form of average vehicle miles traveled per ZIP code. In order to get it there, a significant and time consuming transformation must **occur** The first step is to determine which vehicles are likely to be on the road in 2008, the study year. The first step to do this is to crawl the CARFAX Vehicle History Database to pull out vehicle identification numbers (VINS) from the desired model years (MY). Ideally we would pull VINS from all model years going back to 1995 where the database contains such data. However, due to the immense amount of time and computational power that the process takes, that method would not yield a result for approximately 34 weeks, far beyond the acceptable timeline for this project. Instead we pulled a sampling of MY's using an interval of 3 years: 2008, 2005, 2002, 1999. The processing took about a month per model year, though after the first, the programming setup time was be significantly shorter, so the data came a little faster.

The next step is to trim out vehicles that we don't want to include. We only want personal vehicles, so heavy trucks, emergency vehicles, semi –tractors, school buses, recreational vehicles, and motorcycles will be eliminated. We will also eliminate any vehicles that have been Salvage/ Junked, have had their odometer rolled back, have any inconsistencies in their record, or any other history event that indicates the vehicle was not on the road during 2008, the study year.

The majority of that processing took place at CARFAX, by a team lead by Michael Pierce. I received from him a data set with VIN #, Owner #, ZIP code, Ownership dates, and any odometer readings they have for the vehicle.

Miles traveled for each car for the sample year of 2008 is calculated using miles driven between most recent two (or more where available) odometer readings, and the amount of time between the readings. If, for example, a car drove 100 miles over 3 months in 2008, its 2008 VMT would be calculated as 600 miles. Older odometer readings for the car within the same ZIP code are also used to extrapolate forward to 2008 and add accuracy to the VMT readings.

The VMT's for all cars within a ZIP code are then averaged, yielding the Average VMT per ZIP code. That number is then multiplied by the total number of cars registered in that ZIP. This gives a pool total, or the approximate total number of miles that were driven by cars in that ZIP. A company called POLK provided the data on the number of vehicles currently operating in each ZIP code. POLK was generous enough to lend this data to me for the sake of this project; without this support for academic research, the project could not have happened, as the cost of this data was well beyond the project budget.

There were some relatively major data hiccups at this point. Looking at the distribution of average yearly milage per car, we found a typical normal distribution, but with a steep secondary peak at 0 miles per year. (Figure 11: Distribution of VMT) After exploring the issue, we came to the conclusion that the majority of these 0's were caused by vehicles sitting on dealer lots, and other vehicles not actively in service. For this reason the cars with an average of 0 MPY were thrown out. This methodology was possibly questionable, because there was no way of knowing if the way POLK calculated number of private passenger cars per ZIP code included these dealer and/or inactive vehicles, but all signs indicated that they would not.

In order to tie it back to what matters, this total number of miles traveled is divided by the population of the ZIP code, resulting in the end variable of VMT per person per ZIP code. VMT per person is what matters to this study because I am interested in looking at what factors affect how much people drive in their daily lives. If I were to just use the variable of average VMT per car per zip code, what I would be looking at is travel intensity not total vehicular use. All I would know is that the cars that exist in that ZIP drive more, it could be that there are just very few cars per person – lots of big family for example – not that the structure of the area is particularly conducive to driving. Alternatively, many past studies on structural effects on VMT have used VMT per household as their dependent variable. I believe that this was done because their assumptions were that a household was the basic unit of car use. Essentially I believe that using households as the basic unit assumes that there was one commuter per household, and the other car use was dependent on the activity of the stay-at-home parent. It seems to me that it is easier and better to use individuals as the most basic unit. One problem with this approach is that under the age of 16, people are not drivers, so it has been suggested that the unit should be VMT per person of driving age. While this analysis could be interesting, I think that a better way to separate out and account for the effects of having children is on VMT is to include age breakdown variables in the regression. This way if the model will show if how much the proportion of the population that falls into the younger than 15 age bracket effects VMT.

ESRI is the largest research and development organizations dedicate to Geographic Information Systems (GIS), a field it created in the 1980's. It is an industry leader in geographic data services, and the producer of the leading ArcGIS family for software products. Part of it's business is collecting and producing data for use with GIS software. ESRI keep up to date a number of data sources, they produce a yearly overview data set that includes: Business Data —Business Locations and Business Summaries, Bank Branches, Banking Potential, Cable Boundaries, Retail MarketPlace, Major Shopping Centers, and Traffic Counts databases; Consumer spending data — total expenditures and average spending by household, and Market Potential data on demand for products and services; and Demographic data—Updated demographics including current-year estimates and five-year projections; Tapestry Segmentation, crime indexes, and census data

Demographic Data:

The updated demographics offered by ESRI are sold as their most accurate demographic data for the United States. They include over 2,000 data variables at current year estimates and at 5-year projections. The household characteristics include population, age, sex, race, marital status, educational attainment and household type. Population by sex and age also include estimates of age groups from 1 year to 84 years.

The population by age and sex is based off 2000 data, and project forward via a "cohort survival model that calculates the components of population change separately, by age and sex" (ESRI, 2009) through the application of survival rates for each specific cohort.

The data on racial breakdowns are based on the 2007 Census Bureau's estimates. Survey data is brought to the current year in conjunction with the known 1990 to 2000 change rates. The forecasts are done by block group as well to combine local changes in distribution and overall projected changes at the county level. The Diversity Index from ESRI represents the likelihood that two persons, chosen at random from the same area, belong to different races or ethnic groups. Ethnic diversity, as well as racial diversity, is included in our definition of the Diversity Index. ESRI counts seven race group, one of which is two-or-more races. Each of these groups is divided by hispanic or non-hispanic origin. Scores on the diversity index go from 0 to 100. 0 mean no diversity, and 100 is complete diversity, or a perfect division of race between all groups. The average diversity score for the United States in 2009 was 61, with some states as low as 9.4 and others as high as 83. (ESRI, 2009) A score of 61 means that by taking any two people at random from the U.S. population, there is a 61% chance they will belong to different racial groups.

Average household sizes are very stable, so easy to predict. Nationally they have stabilized at 2.59 persons per household and have been around that number since the 1990s. The ESRI numbers are projections based on census figures.

The breakdown of owner versus renter occupied housing are based on data from the Housing Vacancy Survey, the Current Population Survey, and guided by the latest census data. This method of forecasting gains accuracy by using both top down and bottoms-up techniques. Both survey data at high geographic levels, and local characteristics are taken into account. Changes in owner versus renter occupancy are forecast separately, and sum to the total households.

Home value estimates are based on the House Price Index (HPI) from the Office of Federal Housing Enterprise Oversight (OFHEO). The index is derived from mortgage loans purchase or secured by Freddie Mac or Fannie Mae. This is a significantly better way of collecting this data versus the Commerce Department surveys, and is especially effective when using the data at a smaller geographic area, such as ZIP codes.

Labor force estimates are based on the Local Area Unemployment Statistics (LAUS), Employment Projections (EP), Occupational Employment Statistics (OES), and Current Employment Statistics program of the Bureau of Labor Statistics. They are also based on the American Community Survey (ACS) produced by the Census Bureau. This, again, uses both the top down and bottoms up methodology to optimal accuracy. Employment and unemployment statistics are forecast based on census 2000 and updated based on trends adapted from LAUS.

Industry and occupation statistics were updated using trends from CES. However, because CES excludes farm employment, Current Population Survey (CPS) and ACS were used to fill in the holes.

The projections ESRI reported for income come from the census 2000. After 2000, ESRI evaluated income data from the CPS and ACS and found that they varied from what was reported in 2000. For this reason it was concluded that one point in time data was not a good estimate of the true population value. Estimating post 2000 income levels, then, ESRI relied heavily on time series' data that had proven accurate throughout the 1990's. Yearly updates evaluate current wage inflation and other economic shocks that impact income growth.

The Census data comes formatted in such a way that one row of the table represents one NACIS sub-category for one ZIP code, and each column represented the number of businesses in a number of size groups. This resulted in an unwieldy table of many millions of rows. In order to simplify the table and make it useful, I first summed all sub-categories into industry totals. In order to transform the business size data into a usable form I multiplied the number of business in each size group, and multiplied that by the mean value for that group, i.e. if there were 5 construction companies sized 100 - 200 people, the equation would have been (5 * (average: 100, 200)) = 750 jobs. For the size group of 1000+ I used the value of 2000 for the average, because experimentation with the data showed me that 2000 people was the average size of business over 1000 people. While I believe this method successfully captures the approximate number of jobs in each industry for each ZIP code, there is the potential that huge businesses will be under represented. However the way the census defines a business is by a single location, so if, for example a company had 5 buildings with 5 separate addresses, all 5 would be counted separately. The danger to the accuracy of the system then only comes from huge single building offices. The end result of the transformation was my desired variable or number of jobs in each industry

In order to simplify the industry definitions for my analysis, the 20 categories were grouped into 7 categories: Consumer hub, Office jobs, Factory jobs, Construction, Movers of goods, and Other. The assignment of NACIS categories into these groupings is shown in table (NTA). These groupings are meant not to reflect what the business produce or to group them with similar industries. Instead these grouping are meant to created categories corresponding to the driving demand for the business. Construction jobs then, are likely to require more driving then office jobs, because the job site is more unpredictable and thus housing cannot be picked based off job proximity. Consumer hub jobs, represent those jobs that require the consumer to travel as well, such as retail jobs, schools, hospitals, etc. These jobs are expected to contribute more to VMT because they involved not only the commute of the employee, but the distance the consumers need to drive to reach the service as well. In order to separate out

The census data was transformed such that we had a table for each ZIP code that had columns for “total jobs per capita,” “Consumer hub jobs per capita,” “Office jobs per capita,” “factory jobs per capita,” “Construction jobs per capita,” “Movers of goods jobs per capita,” and “other jobs per capita." These categories were obtained by summing all NACIS categories and subcategories, then dividing these groupings by the population of the ZIP code they describe.

Rings

One of the most original and challenging aspects of this analysis was the inclusion of the business location data from surrounding ZIP codes into the analysis. The first step to this analysis, was producing lists for each ZIP code of the ZIP codes within 5, 10, 20, and 30 miles. The distance between ZIPS was calculated in ArcView using the geographic centroids for each ZIP code. Once this was established, producing the optimal formula to answer the question of 'what is the effect of the pull of job opportunities at the sequential distances from the ZIP code on VMT?' The formula used was:

Jobs[ring]

pop 16+ [home zip]

The rings were also grouped together to concentrate their predicative power. Variables were made that represented the sum of rings less than or equal to five miles, and rings between 5 and 30 miles. These consolidated rings were found to simplify and magnify the effects, and were therefore included in the complete regression instead of the individual rings.

Rings were produced that represented both the numbers of jobs in each industry metacategory, and the total number of jobs. Because of high collinearity between industries, the rings that represented the total jobs per capita were found to be the most effective predictors.

Appendix D **Variable list: All Variables Considered**

VMT_per_Capita
Construction_jobs_perCap <= 5 miles
Factory_jobs_perCap <= 5 miles
Movement_of_goods_jobs_perCap <= 5 miles
Other_jobs_perCap <= 5 miles
16+ Population
Employed
2009 Total Population
% of Population under 15
% of Pop 15-64
% of Pop 65+
Pct_Pop_15_Plus
% male
% White
% Black
% Asian
% Hispanic
% Native American
% other race
% 2 or more races
2009 % of people in families
% of People Who Own Cars
ZIP with more than 50% of Residents in Group Quarters
ZIP with 50-80% of Residents in Group Quarters
ZIP with more than 80% of Residents in Group Quarters
2009 % of people group quarters
2009 Population Density
2009 Population Density (Binned)
2009 Diversity Index
2009 Average Household Size
2009 Average Family Size
2009 Median Household Income
Log of Median Household Income
2009 Per Capita Income
2009 % Housing Units Owner Occupied
2009 % Housing Units Renter Occupied
2009 % Housing Units Vacant
2009 Owner Occupied Housing Unit per Renter Occupied Housing Unit
2009 Median Value of Owner Occupied Housing Units
2000-2009 Population: Annual Compound Growth Rate
2000-2009 Median Household Income: Annual Compound Growth Rate
2000-2009 Per Capita Income: Annual Compound Growth Rate
Pct_pop_in_GroupQuarters<.99 (FILTER)
Metropolitan Statistical Area + Primary MSA
Train trips per Year (unlinked)
Train Transit Available
Bus trips per Year (unlinked)
Bus Transit Available
Other trips per Year (unlinked)
Total trips per Year (unlinked)
Any Public Transit Available
Train + Bus Trips per Year
% of 2009 Population 25+ with a High School Diploma or Higher
% of 2009 Population 25+ with a Bachelor's Degree or Higher
Total jobs
Census total
Census Consumer hub
Census office jobs
Census construction
Census factory jobs
Census movemnt of goods
Census other
Workforce total
unemployed pop
Unemployment Rate
esri consumer hub
esri office jobs
esri construction
esri factory jobs
esri movement of goods

esri other jobs
% of Drivers Who Own Cars
High Population Density Flag
Medium Population Density Flag
Low Population Density Flag
Total Jobs per Cap <= 5 miles
Total Jobs per Cap > 5 miles away
Consumer_hub_jobs_perCap > 5 miles away
Office_jobs_perCap > 5 miles away
Construction_jobs_perCap > 5 miles away
Factory_jobs_perCap > 5 miles away
Movement_of_goods_jobs_perCap > 5 miles away
Other_jobs_perCap > 5 miles away
Retail_Trade_jobs_sum > 5 miles away
Retail_Trade_jobs_sum <= 5 miles
Consumer_hub_jobs_perCap <= 5 miles
Train trips per Year (unlinked) (by city)
Train Transit Available (by city)
Bus trips per Year (unlinked) (by city)
Bus Transit Available (by city)
Other_new
total_new
total_train__bus

Appendix E — Descriptive Statistics for Final Variables

Descriptive Statistics					
	Min.	Max	Mean	Std. Deviation	Variance
VMT_per_Capita	7.83	69227.32	12874.8	5973.98622	3.57E+07
2009 Total Population	1	116577	10414.1	14156.52784	2.00E+08
% of Pop 15-64	14.9	100.1	66.3861	4.97188	24.72
% male	5.4	100	49.989	3.2766	10.736
2009 Median Household Income	1	348129	50481.7	21043.4186	4.43E+08
2009 Diversity Index	0	95.9	29.4972	24.30455	590.711
% of 2009 Population 25+ with a Bachelor's Degree or Higher	0	0.95	0.208	0.14002	0.02
2009 % of people in families	0	1	0.8058	0.11139	0.012
2009 % of people group quarters	0	1	0.0294	0.09559	0.009
2009 Average Family Size	0	9	3.0309	0.36856	0.136
2009 Population Density	0	147367.5	1266.25	4763.86479	2.27E+07
2009 % Housing Units Renter Occupied	0	1	0.2207	0.13721	0.019
2009 % Housing Units Vacant	0	0.99	0.156	0.13033	0.017
Unemployment Rate	0	80.3	9.7399	5.00983	25.098
Total Jobs per Cap > 5 miles away	0	78509.32	246.766	1264.88898	1.60E+06
Total Jobs per Cap <= 5 miles	0	5111.58	11.2532	107.4545	11546.469
Train trips per Year (unlinked) (by city)	0	155.04	0.7342	7.17857	51.532
Train Transit Available (by city)	0	1	0.0399	0.19568	0.038
Bus trips per Year (unlinked) (by city)	0	99.9	2.4627	8.85378	78.389
Bus Transit Available (by city)	0	1	0.1215	0.32675	0.107
2000-2009 Population: Annual Compound Growth Rate	-37.5	135.2	0.8094	2.76982	7.672
2000-2009 Median Household Income: Annual Compound Growth Rate	-100	19.9	2.3549	1.24946	1.561

	VMT_per_Capita	% of Pop 15-64	% male	2009 Median Household Income	2009 Diversity Index	% of 2009 Population 25+ with a Bachelor's Degree or Higher	2009 % of people in families	2009 % of people group quarters	2009 Average Family Size	2009 Population Density	2009 % Housing Units Renter Occupied	2009 % Housing Units Vacant	Unemployment Rate	Total Jobs per Cap > 5 <= 30 miles	Total Jobs per Cap <= 5 miles	Train trips per Year (unlinked) (by city)	Train Transit Available (by city)	Bus trips per Year (unlinked) (by city)	Bus Transit Available (by city)	2000-2009 Population: Annual Compound Growth Rate	2000-2009 Median Household Income: Annual Compound Growth Rate
VMT_per_Capita	1.000	-0.156	0.052	-0.105	-0.307	-0.221	0.289	-0.222	-0.172	-0.269	-0.404	0.187	-0.179	0.061	-0.056	-0.133	-0.194	-0.254	-0.287	-0.087	-0.099
% of Pop 15-64	-0.156	1.000	0.364	0.081	0.187	0.260	-0.562	0.528	-0.040	0.187	0.357	-0.129	0.055	0.088	0.170	0.138	0.161	0.221	0.198	0.074	-0.057
% male	0.052	0.364	1.000	-0.088	0.054	-0.148	-0.230	0.332	0.014	-0.099	-0.007	0.160	-0.022	0.063	0.085	-0.026	-0.022	-0.041	-0.076	0.024	-0.075
2009 Median Household Income	-0.105	0.081	-0.088	1.000	0.043	0.718	0.164	-0.112	0.135	0.083	-0.084	-0.324	-0.237	0.128	0.038	0.062	0.051	0.056	0.022	0.198	0.240
2009 Diversity Index	-0.307	0.187	0.054	0.043	1.000	0.147	-0.241	0.171	0.345	0.300	0.538	-0.136	0.294	0.021	0.090	0.128	0.254	0.304	0.342	0.138	0.060
% of 2009 Population 25+ with a Bachelor's Degree or	-0.221	0.260	-0.148	0.718	0.147	1.000	-0.265	0.115	-0.060	0.224	0.224	-0.164	-0.221	0.113	0.121	0.166	0.176	0.249	0.229	0.132	0.153
2009 % of people in families	0.289	-0.562	-0.230	0.164	-0.241	-0.265	1.000	-0.793	0.255	-0.250	-0.578	-0.078	-0.144	-0.112	-0.265	-0.184	-0.228	-0.316	-0.337	0.016	0.020
2009 % of people group quarters	-0.222	0.528	0.332	-0.112	0.171	0.115	-0.793	1.000	-0.114	0.096	0.345	-0.007	0.106	0.148	0.246	0.033	0.061	0.103	0.135	0.024	-0.113
2009 Average Family Size	-0.172	-0.040	0.014	0.135	0.345	-0.060	0.255	-0.114	1.000	0.110	0.212	-0.271	0.219	-0.030	-0.064	0.006	0.083	0.068	0.071	0.076	0.109
2009 Population Density	-0.269	0.187	-0.099	0.083	0.300	0.224	-0.250	0.096	0.110	1.000	0.484	-0.131	0.142	0.015	0.082	0.614	0.353	0.434	0.275	0.000	0.084
2009 % Housing Units Renter Occupied	-0.404	0.357	-0.007	-0.084	0.538	0.224	-0.578	0.345	0.212	0.484	1.000	-0.306	0.278	0.036	0.191	0.252	0.344	0.442	0.467	-0.028	0.076
2009 % Housing Units Vacant	0.187	-0.129	0.160	-0.324	-0.136	-0.164	-0.078	-0.007	-0.271	-0.131	-0.306	1.000	0.033	-0.037	-0.026	-0.041	-0.081	-0.121	-0.138	-0.094	-0.082
Unemployment Rate	-0.179	0.055	-0.022	-0.237	0.294	-0.221	-0.144	0.106	0.219	0.142	0.278	0.033	1.000	-0.016	0.029	0.074	0.128	0.139	0.165	-0.017	-0.031
Total Jobs per Cap > 5 <= 30 miles	0.061	0.088	0.063	0.128	0.021	0.113	-0.112	0.148	-0.030	0.015	0.036	-0.037	-0.016	1.000	0.536	0.028	0.030	0.021	0.001	0.033	0.014
Total Jobs per Cap <= 5 miles	-0.056	0.170	0.085	0.038	0.090	0.121	-0.265	0.246	-0.064	0.082	0.191	-0.026	0.029	0.536	1.000	0.140	0.130	0.144	0.125	0.046	0.032
Train trips per Year (unlinked) (by city)	-0.133	0.138	-0.026	0.062	0.128	0.166	-0.184	0.033	0.006	0.614	0.252	-0.041	0.074	0.028	0.140	1.000	0.502	0.558	0.275	0.013	0.071
Train Transit Available (by city)	-0.194	0.161	-0.022	0.051	0.254	0.176	-0.228	0.061	0.083	0.353	0.344	-0.081	0.128	0.030	0.130	0.502	1.000	0.696	0.548	0.039	0.074
Bus trips per Year (unlinked) (by city)	-0.254	0.221	-0.041	0.056	0.304	0.249	-0.316	0.103	0.068	0.434	0.442	-0.121	0.139	0.021	0.144	0.558	0.696	1.000	0.748	0.069	0.073
Bus Transit Available (by city)	-0.287	0.198	-0.076	0.022	0.342	0.229	-0.337	0.135	0.071	0.275	0.467	-0.138	0.165	0.001	0.125	0.275	0.548	0.748	1.000	0.064	0.062
2000-2009 Population: Annual Compound Growth Rate	-0.087	0.074	0.024	0.198	0.138	0.132	0.016	0.024	0.076	0.000	-0.028	-0.094	-0.017	0.033	0.046	0.013	0.039	0.069	0.064	1.000	0.028
2000-2009 Median Household Income: Annual	-0.099	-0.057	-0.075	0.240	0.060	0.153	0.020	-0.113	0.109	0.084	0.076	-0.082	-0.031	0.014	0.032	0.071	0.074	0.073	0.062	0.028	1.000

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.290a	0.084	0.084	5718.00721

Coefficientsa					
Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	14.359	420.724		0.034	0.973
2009 % of people in families	162.62	490.106	0.303	33.183	0
2009 % of people not in families	-16.6	467.594	-0.02	-3.551	0
2009 % of people group quarters	11.25	570.649	0.018	1.973	0.049

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.289a	0.084	0.084	5719.46048
2	.030c	0.001	0.001	5972.57435
3	.222d	0.049	0.049	5825.67453

Coefficientsa

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	362.663	242.722		1.494	0.135
	2009 % of people in families	15531.408	298.394	0.289	52.05	0
2	(Constant)	13288.485	87.744		151.447	0
	2009 % of people not in families	-2487.854	487.788	-0.03	-5.1	0
3	(Constant)	13286.352	35.408		375.232	0
	2009 % of people group quarters	-13897.592	354.071	-0.222	-39.251	0

Model Summary									
Model	R	R^2	Adjusted R^2	Std. Error of the Estimate	Change Statistics				
					R^2 Change	F Change	df1	df2	Sig. F
1	.350a	0.122	0.122	5200.50766	0.122	679.919	6	29242	0
2	.400c	0.16	0.16	5087.46653	0.16	929.457	6	29242	0
3	.122d	0.015	0.015	5509.5043	0.015	221.348	2	29246	0
4	.408e	0.166	0.166	5069.21669	0.166	416.85	14	29234	0
5	.418f	0.175	0.174	5043.62069	0.175	258.448	24	29224	0

Coefficientsa								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	13833	35.245		392.49	0		
	Census Consumer hub	-0.598	0.016	-0.302	-37.85	0	0.47	2.125
	Census office jobs	0.019	0.012	0.012	1.604	0.109	0.497	2.011
	Census construction	-0.554	0.064	-0.064	-8.633	0	0.54	1.851
	Census factory jobs	-0.17	0.035	-0.036	-4.882	0	0.552	1.812
	Census movemnt of goods	-0.023	0.042	-0.005	-0.558	0.577	0.432	2.316
	Census other	0.277	0.134	0.013	2.07	0.038	0.791	1.264
2	(Constant)	14346	37.702		380.51	0		
	esri consumer hub	-0.791	0.033	-0.418	-24.34	0	0.098	10.247
	esri office jobs	-0.016	0.047	-0.004	-0.35	0.727	0.198	5.047
	esri construction	0.388	0.137	0.031	2.825	0.005	0.232	4.302
	esri factory jobs	0.137	0.072	0.017	1.893	0.058	0.377	2.652

	esri movement of goods	0.691	0.138	0.065	5.008	0	0.17	5.885
	esri other jobs	-1.249	0.217	-0.08	-5.751	0	0.147	6.816
3	(Constant)	12676	32.813		386.31	0		
	Total Jobs per Cap <= 5 miles	-6.464	0.354	-0.126	-18.25	0	0.712	1.405
	Total Jobs per Cap > 5 miles away	0.561	0.03	0.128	18.628	0	0.712	1.405
4	(Constant)	13272	35.698		371.79	0		
	Consumer_hub_jobs_perCap > 5 miles away	3.783	0.578	0.3	6.544	0	0.014	73.503
	Office_jobs_perCap > 5 miles away	-6.56	0.434	-0.487	-15.12	0	0.027	36.418
	Construction_jobs_perCap > 5 miles away	28.92	2.028	0.379	14.263	0	0.04	24.759
	Factory_jobs_perCap > 5 miles away	15.068	1.151	0.286	13.092	0	0.06	16.698
	Movement_of_goods_jobs_perCap > 5 miles away	-16.36	1.541	-0.346	-10.62	0	0.027	37.307
	Other_jobs_perCap > 5 miles away	11.997	2.728	0.083	4.398	0	0.081	12.405
	Retail_Trade_jobs_sum > 5 miles away	-0.007	0	-0.143	-19.47	0	0.53	1.887
	Retail_Trade_jobs_sum <= 5 miles	-0.107	0.004	-0.22	-29.29	0	0.507	1.972
	Consumer_hub_jobs_perCap <= 5 miles	-66.11	3.785	-0.393	-17.47	0	0.056	17.789
	Office_jobs_perCap <= 5 miles	25.405	1.934	0.187	13.133	0	0.141	7.112
	Construction_jobs_perCap <= 5 miles	54.901	14.814	0.055	3.706	0	0.131	7.662
	Factory_jobs_perCap <= 5 miles	-48.74	8.743	-0.066	-5.574	0	0.203	4.927
	Movement_of_goods_jobs_perCap <= 5 miles	45.968	8.11	0.082	5.668	0	0.136	7.35
	Other_jobs_perCap <= 5 miles	261.86	22.69	0.106	11.541	0	0.338	2.957
5	(Constant)	13064	36.32		359.69	0		
	Retail_Trade_jobs_sum 5 - 10 miles	-0.074	0.003	-0.285	-28.28	0	0.279	3.59
	Retail_Trade_jobs_sum 10 - 20 miles	-0.002	0.001	-0.016	-1.232	0.218	0.163	6.139
	Retail_Trade_jobs_sum 20 - 30 miles	-0.003	0.001	-0.029	-3.012	0.003	0.312	3.206
	Total_Jobs_perCap 5 - 10 miles	-8.883	4.064	-0.28	-2.186	0.029	0.002	580.6
	Total_Jobs_perCap 10 - 20 miles	25.415	2.445	2.302	10.397	0	0.001	1737.3

	Total_Jobs_perCap 20 - 30 miles	18.812	1.571	2.375	11.977	0	0.001	1393.5
	Consumer_hub_jobs_perCap 5 - 10 miles	10.747	6.178	0.105	1.74	0.082	0.008	130.18
	Consumer_hub_jobs_perCap 10 - 20 miles	-21.56	3.38	-0.679	-6.379	0	0.002	401.34
	Consumer_hub_jobs_perCap 20 - 30 miles	-17.12	2.244	-0.762	-7.628	0	0.003	353.23
	Office_jobs_perCap 5 - 10 miles	19.419	5.312	0.209	3.655	0	0.009	116.34
	Office_jobs_perCap 10 - 20 miles	-37.96	2.94	-1.125	-12.91	0	0.004	269.21
	Office_jobs_perCap 20 - 30 miles	-29.44	1.946	-1.245	-15.12	0	0.004	240.11
	Construction_jobs_perCap 5 - 10 miles	-37.46	10.522	-0.067	-3.56	0	0.08	12.568
	Construction_jobs_perCap 10 - 20 miles	-54.49	6.651	-0.296	-8.193	0	0.022	46.259
	Construction_jobs_perCap 20 - 30 miles	28.446	4.322	0.192	6.581	0	0.033	30.167
	Factory_jobs_perCap 5 - 10 miles	-27.89	7.283	-0.071	-3.83	0	0.083	12.009
	Factory_jobs_perCap 10 - 20 miles	-20.62	3.676	-0.151	-5.608	0	0.039	25.559
	Factory_jobs_perCap 20 - 30 miles	2.237	2.721	0.024	0.822	0.411	0.034	29.684
	Movement_of_goods_jobs_perCap 5 - 10 miles	2.201	8.791	0.006	0.25	0.802	0.045	22.252
	Movement_of_goods_jobs_perCap 10 - 20 miles	-27.51	4.622	-0.225	-5.951	0	0.02	50.853
	Movement_of_goods_jobs_perCap 20 - 30 miles	-50.85	2.864	-0.603	-17.75	0	0.024	40.893
	Other_jobs_perCap 5 - 10 miles	111.09	21.854	0.073	5.083	0	0.138	7.263
	Other_jobs_perCap 10 - 20 miles	30.932	9.088	0.087	3.403	0.001	0.044	22.94
	Other_jobs_perCap 20 - 30 miles	54.159	6.333	0.212	8.552	0	0.046	21.755

The variables ultimately chosen to represent job location for the final stage of regressions of this project were extremely simplified. Combining all the industry types into total jobs removed the information on how the type of job affects VMT. Methods to explore in more detail the effects of job type, location, and mix were ultimately rejected as they proved too difficult to interpret high and low R values, posed extra danger of multicollinearity, and did not add enough to the model to make them worthwhile.

Two of these methods involved using data that directly described each ZIP code: one used census data on the location and size of places of work by industry, and the other used ESRI data on the number of workers in each industry, by zip code. The first set showed where the jobs are, the second showed where the workers live. One problem with this methodology is that many of the industry grouping variables used were not significant when modeled against VMT.

The ring variables described earlier were also explored for their predicative power. The less than 5, 10, 20, and 30 mile rings produced results too inconsistent, and insignificant to draw any major conclusions. The summarized rings which combined the 10, 20, and 30 miles rings into one, retained much of the predictive power but simplified the effects. However, there was still too much multicollinearity between the industry categories, so the third option was chosen; only two ring variables were used, total jobs less than or equal to five miles, and total jobs greater than five miles but less than 30 miles.

Model summery				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.307a	0.094	0.094	5686.2452
3	.019c	0	0	5973.51564
4	.308d	0.095	0.094	5684.76651

Coefficients								
Model		Unstandardized Coefficients		Standardized C	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	15098.18	51.854		291.2	0		
	2009 Diversity Index	-75.377	1.357	-0.307	-55.56	0	1	1
	(Constant)	12912	50.71		254.6	0		
	% Black	2.04	2.112	0.006	0.966	0.334	0.993	1.007
	% Asian	3.332	8.071	0.003	0.413	0.68	0.701	1.427
	% Hispanic	3.852	5.934	0.009	0.649	0.516	0.157	6.369
	% Native American	-10.631	4.333	-0.015	-2.453	0.014	0.957	1.045
	% other race	-7.074	12.204	-0.008	-0.58	0.562	0.16	6.244
	% 2 or more races	-24.82	20.204	-0.009	-1.228	0.219	0.595	1.681
4	(Constant)	15168.97	63.033		240.7	0		
	2009 Diversity Index	-75.536	1.357	-0.307	-55.66	0	0.999	1.001
	% Black	1.33	2.009	0.004	0.662	0.508	0.993	1.007
	% Asian	-5.154	7.682	-0.004	-0.671	0.502	0.701	1.428
	% Hispanic	1.228	5.647	0.003	0.217	0.828	0.157	6.369
	% Native American	-16.364	4.125	-0.022	-3.967	0	0.956	1.046
	% other race	-4.282	11.614	-0.005	-0.369	0.712	0.16	6.245
	% 2 or more races	-16.818	19.228	-0.006	-0.875	0.382	0.595	1.681