REDUCING VEHICLE MILES TRAVELED THROUGH SMART LAND-USE DESIGN

FINAL REPORT C-08-29

Prepared for the

NEW YORK STATE ENERGY RESEARCH AND DEVELOPMENT AUTHORITY
Albany, NY
Joseph D. Tario
Senior Project Manager
&
NEW YORK STATE DEPARTMENT OF TRANSPORTATION
Albany, NY
Gary Frederick
Director, Research and Development Bureau

Prepared by

Adel W. Sadek, Ph.D., Associate Professor[1]

Qian Wang, Ph.D., Assistant Professor[1]

Peng Su, Graduate Research Assistant[1]

Andrew Tracy, Graduate Research Assistant[1]

[1] Department of Civil, Structural and Environmental Engineering
University at Buffalo, the State University of New York

NYSERDA Contract No. 11107

December 2011

NOTICE

DISCLAIMER

1. Report No. C-08-29	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Reducing Vehicle Miles Traveled through Smart Land-use Design		5. Report Date December 2011
		6. Performing Organization Code
7. Author(s) A.W. Sadek, Q. Wang, P. Su, A.J. Tracy		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil, Structural and Environmental Engineering University at Buffalo, the State University of New York 212 Ketter Hall, Buffalo, NY 14260		10. Work Unit No.
		11. Contract or Grant No. NYSERDA Contract No. 11107
12. Sponsoring Agency Name and Address New York State Energy Research and Development Authority (NYSERDA), 17 Columbia Circle, Albany, NY 12203; New York State Department of Transportation (NYSDOT), 50 Wolf Road, Albany, NY 12232		13. Type of Report and Period Covered Final Report (2009 – 2011)
		14. Sponsoring Agency Code

15. Supplementary Notes
The project was funded in part with funds from the Federal Highway Administration. Joseph D. Tario from NYSERDA and Gary Frederick from NYSDOT served as project managers.

16. Abstract
The objective of the present study was to help planning and transportation organizations across New York State identify the most appropriate methods that can evaluate the likely impacts of smart growth strategies. Two approaches were investigated using the Greater Buffalo/Niagara metropolitan area as a case study. The first approach involved a GIS-based methodology by which spatial characteristics of the built environment were quantified and used to predict travel behavior at an aggregate level. A wide scope of travel behavior was examined, and over 50 variables, many of which are based on high-detail data sources, were investigated for potentially quantifying the built environment. Linear modeling was then used to relate travel behavior and the built environment, yielding models that may be applied in a post-processor fashion to travel models' results to provide some measure of sensitivity to built environment modifications. The second approach developed an enhanced travel demand forecasting method to evaluate the impact of smart growth strategies on travel patterns. Though the modelling framework shares a similar structure as the traditional four-step planning method, behavior choice models were developed in order to capture the impact of land use on travel behavior. The enhanced travel demand forecasting method was tested using the Greater Buffalo-Niagara Area as the study case. Findings support the claims that compact, mixed-use, pedestrian-friendly and transit-friendly designs can reduce vehicle trips, encourage non-motorized modes, decrease average trip length, and reduce daily VMT. Moreover, the study has developed two useful methodologies which can be applied to increase the sensitivity of current modeling tools toward assessing the likely impacts of proposed smart growth strategies.

17. Key Words Smart growth, sustainable transportation, built environment, four-step process, post-processor models, ,land-use	18. Distribution Statement No Restrictions		
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 232	22. Price

Form DOT F 1700.7 (8-72)

EXECUTIVE SUMMARY

Recent research tends to suggest that smart growth may be one strategy which can help reduce Vehicle Miles Traveled (VMT) and improve the overall sustainability of the transportation system. Several other studies have concluded that mixed and diverse land use development can be effective in reducing urban sprawl, shortening the length of vehicle trips, encouraging the use of transit, and making walking and bicycling possible. As a result, there has been an increased interest among planning organizations, the federal and state government in developing and implementing smart growth strategies. The development and implementing of smart growth strategies, however, requires modeling tools that are sensitive enough to reflect their likely benefits. This is unfortunately not the case with current travel demand modeling practice centered on the ubiquitous four-step planning process(Ben-Akiva and Lerman 1985).

This project was designed to first help planning and transportation organizations across New York State identify the most appropriate methods that would allow for better reflecting the benefits of smart growth in travel demand forecasting practice. Two approaches for increasing the sensitivity of transportation models to smart growth impact were then investigated in greater detail, using the Greater Buffalo/Niagara metropolitan area as a case study. The first approach involved a GIS-based methodology by which spatial characteristics of the built environment were quantified and used to predict travel behavior at an aggregate level. A wide scope of travel behavior was examined, and over 50 variables, many of which are based on high-detail data sources, were investigated for potentially quantifying the built environment. Linear modeling was then used to relate travel behavior and the built environment, yielding models that may be applied in a post-processor fashion to travel models to provide some measure of sensitivity to built environment modifications.

With respect to the first approach, the modeling exercise revealed that zonal mode choice is highly correlated to built environment factors, while household home-based VHT and VMT are less so. Statistical concepts such as regression C_p minimization, principal component analysis, and power transformations were explored and found to be methodologically beneficial. To conclude the study, the method was applied to a hypothetical land use scenario to estimate the reduction in zonal vehicle dependency caused by high-density development in suburban areas.

The second approach, on the other hand, developed an enhanced travel demand forecasting method to evaluate the impact of smart growth strategies on travel patterns. Though the modelling framework shares a similar structure as the traditional four-step planning method, behavior choice models were developed in order to capture the impact of land use on individual travellers' various travel decisions such as intrazonal trip making, destination choices and mode choices. The enhanced travel demand forecasting method was tested by using the Greater Buffalo-Niagara Area as the study case. As found, population density and employment density exert some influence on mode choice by increasing the utilities of walk and bike. Diversified land uses either encourage walk or discourage automobile in intrazonal trips. Interzonal trips are less influenced by land use variables than intrazonal trips. Destination TAZ's population density and employment density are positively associated with usage of walk. Another finding is about how people's destination choices are correlated with land use patterns. The diversity of land uses and transit-oriented designs play an important role in reducing the average trip length and Vehicle-Miles-Travelled (VMT). Overall, the elasticity of VMT and land use entropy of all the TAZs is -0.027, which means a 10% increase of land use entropy in this area will cause a 0.27% reduction of daily VMT.

The study's findings support the claims that compact, mixed-use, pedestrian-friendly and transit-friendly designs can reduce vehicle trips, encourage non-motorized modes, decrease average trip length, and reduce daily VMT. Moreover, the research has resulted in development of two useful methodologies which can be applied to increase the sensitivity of current modeling tools toward assessing the likely impacts of proposed smart growth strategies.

ACKNOWLEDGMENTS

The members of the research team gratefully acknowledge sponsorship of this project by the New York State Energy Research and Development Authority (NYSERDA) and the New York State Department of Transportation (NYSDOT), under the direction of Joseph D. Tario of NYSERDA and Gary Frederick of NYSDOT. We also acknowledge the very helpful input and support throughout the course of the project provided by the Greater Buffalo-Niagara Regional Transportation Council (GBNRTC) under the direction of Hal Morse, GBNRTC Executive Director.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1 Background

In 1903, Henry Ford's revolutionary production line techniques made automobiles available to most people in the US. For the first in the history, workers in a factory could afford the product they manufactured. The vast production of automobiles and following large scale construction of freeway systems changed the shape of the country. The unprecedented mobility provided by cheap automobiles and the Interstate Highway System made it possible for workers to live tens of miles away from their work place. The isolation of land-use functions (i.e. residential, commercial, employment), leads to a development pattern dominated by housing-only enclaves, distanced trip origins and destinations, and low population densities. They result in continued increase in Vehicle-Miles-Travelled (VMT), more congested roads, increased energy use, deteriorating air quality, and increased emissions of greenhouse gases. In one word, such a development pattern is very unsustainable and naturally unfavourable.

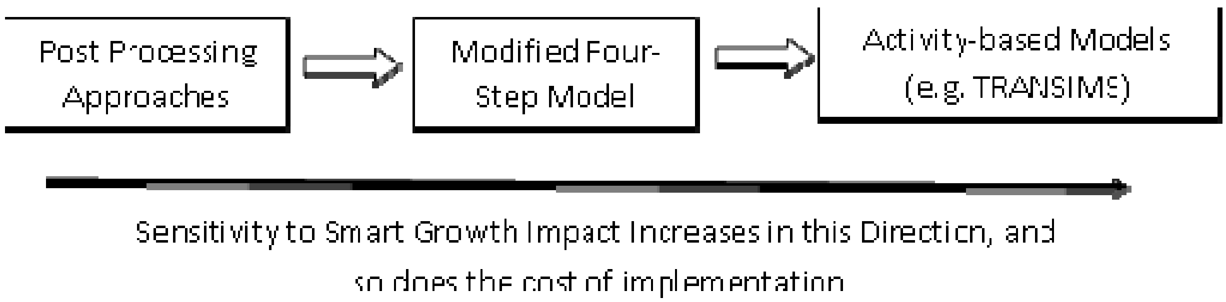
Recently, there has been a renewed interest in better understanding and designing the land use and transportation system (LUTS). This interest is not only motivated by the need to relieve congestion, but, more importantly, by the increased national interest in environmental protection and sustainability, and in light of current woes about future energy shortage. Among land-use strategies that are being currently investigated for their potential in reducing VMT and improving transportation system sustainability are strategies such as smart growth and neo-traditional neighborhood designs; here "traditional" means the low density and land use function isolation pattern predominant in the US after WWII. The fundamental idea behind such concepts is to revitalize LUTS to replace "sprawl" with more compact and mixed-use communities, such as discouraging dispersed, automobile-dependent development at the urban fringe. One of their ultimate goals is to reduce VMT while maintaining mobility and accessibility of human societies.

This current interest in developing and implementing smart growth strategies points out to the genuine need for methods and modeling tools sensitive enough to reflect the benefits of these strategies in travel demand forecasting and transportation infrastructure planning practice. The reality of the situation, however, is that there are currently very few, *if any*, of such methods that have gained wide acceptance and adoption by MPOs in the country. The current research was motivated by the need to address that gap.

1.2 Scope of the Research

Naturally, there is a wide spectrum of tools and methodologies that could be used to increase the sensitivity of transportation modeling software so as to allow for accurately evaluating the impact of smart growth strategies. These tools will likely vary in terms of their capabilities and sensitivity to smart growth impact, ease of implementation and use, and the resources required to develop and run them. In general, one can categorize the available methodologies or tools for increasing the sensitivity of transportation models to smart growth impact into the following three categories: (1) post-processing approaches that work with a basic four-step transportation planning model; (2) modified implementations of the four-step process that can reflect various aspects of smart growth impact; and (3) disaggregate, activity-based approach. The sensitivity, along with the cost of implementation of such tools varies, as shown in Figure 1.1. A brief description of these three different approaches is given below.

Figure 1.1 Approaches for Increasing Smart Growth Sensitivity of Transportation Models



1.2.1 Post Processing Approaches:

These approaches are based on assessing the impact of the following four D's on reducing vehicle trips and vehicle miles traveled: (1) Density, which refers to population and employment per square mile; (2) diversity, which refers to the ratio of jobs to population; (3) design which pertains to aspects of the pedestrian environment design such as street grid density, sidewalk completeness, and route directness; and (4) Destinations, which refers to accessibility compared to other activity concentrations. Calculating the likely reduction in vehicle trips and VMT is based on elasticity factors such as those documented in (Loudon and Parker 2008). Among the more well-known of the post-processing approaches are two GIS-based programs, INDEX and I-PLACE3S, which have been used in land-use planning exercises to assess or demonstrate the transportation benefits of alternative smart-growth strategies, particularly in California, as well as elsewhere.

Using the Greater Buffalo/Niagara metropolitan area as a case study, this study developed a GIS-based methodology by which spatial characteristics of the built environment were quantified and used to predict travel behavior at an aggregate level. A wide scope of travel behavior was examined, and over 50 variables, many of which are based on high-detail data sources, were investigated for potentially quantifying the built environment. Linear modeling was then used to relate travel behavior and the built environment, yielding models that may be applied in a post-processor fashion to travel models to provide some measure of sensitivity to built environment modifications.

1.2.2 Modified Four Step Models:

The four-step travel forecasting method is by far the most popular planning method currently in use by metropolitan planning organizations to evaluate alternative land use and transportation developments. However, as identified by many researchers, a variety of issues associated with the process limit its applicability to smart growth practices. Specifically, (Loudon and Parker 2008) identified more than ten limitations of the four-step process in relation to smart growth strategy evaluation. These include: (1) no explicit modeling of trip chaining; (2) a focus primarily on vehicle trips only; (3) limited or no modeling capability for transit, walking and bicycling; (4) fixed vehicle trip rates by land-use type regardless of the design of the development; (5) zonal aggregation of traveler characteristics; (6) large traffic analysis zones; and (7) a focus on primarily modeling peak periods. Moreover, (Greenwald 2006) argued that the traditional four-step model processes do not capture the increase in shorter intra-zonal automobile trips, bicycle trips and walking trips that are encouraged by smart-growth strategies, due to the limitations of the four-step process in modeling intra-zonal trips and travel made by means other than automobiles.

To fulfil the needs for methods and models sensitive enough to smart growth strategies, an enhanced travel demand forecasting framework was developed in this research, which offers an increased sensitivity to the impact of smart growth strategies. There were two reasons that motivated the study to focus on to developing the enhanced travel demand forecasting method, in addition to the post-processor methods mentioned above. First, traditional four-step method can be easily adopted by the MPOs due to historical reasons. And secondly, compared with activity-based models, the enhanced travel demand forecasting approach consumes less data in model calibration and validation, thus saves money.

The data used to build the model came from the 2002 Buffalo-Niagara Regional Household Travel Survey. While smart growth policies were not explicitly implemented in this area, some neighbourhoods had higher density and are more mixed-used in natural than others. These can thus be considered as smart growth implementations. Comparing the travel behaviours between traditional TAZs and smart growth implemented TAZs thus helped reveal some insights about the relationship between land use and travel behaviour.

Typically, land use factors are quantitatively denoted by variables which can be categorized into four Ds: Density, Diversity, Design and Destination. The 4Ds variables were incorporated into the enhanced travel demand forecasting model. In this report, land use attributes, built environment and smart growth strategies are three expressions we used interchangeably with each other.

Enhanced Travel Demand Forecasting Framework

The overall framework of the enhanced travel demand forecasting approach is shown in Figure 1.2. Trip generation is set of purpose-specific zone-based linear regression functions. Six trip purposes are considered, including Home-Based-Work (HBW), Home-Based-Shop (HBShop), Home-Based-Social-Recreation (HBSR), Home-Based-Other (HBO), Non-Home-Based-Work (NHBW), and Non-Home-Based-Other (NHBO). Trip generation gives the total number of trips produced by each TAZ, and then trip synthesizer uses some algorithm to generate the socio-economic variables for each of the produced trips.

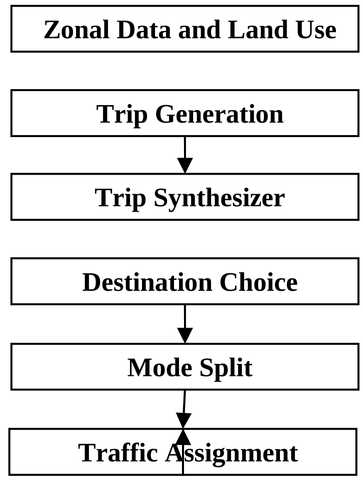


Figure 1.2 Framework of the Enhanced Travel Demand Forecasting Method

In trip distribution, the traditional gravity model is substituted by five trip-based disaggregate destination choice models, including models for HBW, HBShop, HBSRO (combined by HBSR and HBO), NHBW, and NHBO trips. Mode choice models are two nested logit models, one for intrazonal trips and the other for interzonal trips. Both models consider six alternatives modes, including non-motorized modes (walk and bike). Mode choice could split the overall OD table into six sub OD tables, one for each travel mode. In trip assignment, we used All-or-Nothing method for auto OD table, for illustration.

Land use variables are incorporated in three steps: trip generation, destination choice and mode choice. Intrazonal trips are paid special attention to in the enhanced four steps model. As mentioned above, destination choice model takes intrazonal trip as a separate alternative, and intrazonal trips' mode choice is estimated by a separate mode choice model.

1.2.3 Disaggregate Activity-Based Models:

As Figure 1.1shows disaggregate, activity-based models offer the highest level of sensitivity in terms of modeling the impact of smart growth strategies. They are, however, the most complex, and most demanding in terms of implementation cost and effort. In this study, this approach was not investigated in great details because: (1) the approach is very demanding in terms of the data needed for implementation, calibration and validation; and (2) the enhanced four-step modeling framework developed in this study actually captures many of the features and advantages of the activity-based modeling approach. Specifically, many aspects of our modified four-step method are disaggregate in nature and captures individual-level behavior.

1.3 Organization of the Report

In addition to this introduction, this report is organized into five sections or chapters. Section 2 reviews past research efforts related to measuring travel behaviour, quantifying smart growth strategies, and modelling methods for assessing the likely impact of smart growth on travel behaviour. Section 3 is devoted to discussing the first approach pursued in this study to develop smart growth-sensitive modelling tools, namely the one involving developing post-processor models developed in the study to allow for quantifying the likely impacts of smart growth strategies on travel behaviour. Specifically, section 3 describes the modelling methodology, the regression models developed, attempts at improving the accuracy of the developed models through principal component analysis and variable transformation. The section also includes an example to illustrate how the post-processor models may be applied. Section 4 then describes the enhanced four-step travel demand forecasting method developed, including a description of the: (1) the land-use variables used; (2) the trip generation model; (3) the destination choice model; and (4) the mode choice models. Section 4 also includes several scenarios designed to illustrate the effectiveness of the enhanced method in evaluating smart growth strategies. The report concludes with chapter 5 which summarizes the main conclusions of the study. A number of appendices are included which provide more detail about the models developed and the raw data used for the study.

2. LITERATURE REVIEW

The main lessons learned from the literature reviewed during the course of this study will be summarized under the following headings or sections: (1) measures of travel behavior; (2) measures and methods used to characterize the built environment and smart growth strategies; (3) models and methodologies which have been proposed to evaluate the impact of the built environment and smart growth strategies on travel behavior; and (4) the issue of causality in travel behavior models.

2.1Measures of Travel Behavior

VMT reduction is the final goal of all kinds of smart growth strategies. VMT is an integrated product of many factors, such as overall mobility level, vehicle ownership, mode choice, trip lengths, etc. The previous studies used one measure or a combination of measures to reflect the changes in travel behavior.

2.1.1 Overall Mobility Level

The mobility level, or trip generation rates depends primarily on household and personal socio-economic attributes, i.e. income, gender, age, and number of jobs (Ewing and Cervero 2001). Land use attributes could exert very limited influence, if any, on people's mobility level. In a study conducted by Boarnet and Greenwald in 2000 (Boarnet and Greenwald 2000), non-work trip frequency was used as the indicator of travel behavior to build two sets of models, including a census-tract-level model and a zip code-level model. The census-tract-level model could not incorporate any of the land use variables. And the zip code level model includes only one land use variable, which was the density of single-family attached dwelling densities. This variable is associated negatively with non-work auto trip generation at 10% (two-tailed test) significance level.

In a more recent study conducted by Cao et al, the trip frequencies of three modes: auto, transit, and walking/bicycling, were regressed against socio-demographic, attitudinal, and neighborhood characteristics (Cao, Mokhtarian et al. 2009). The results do support the influential power of some design characteristics on non-motorized trip generation rates. But this study is more about mode choice rather than trip generation, and the relationship between land use factors and the overall trip generation rates is not shown in the paper.

2.1.2 Vehicle Ownership

Vehicle ownership is an important indicator of travel patterns, and is a longer-term decision for a family, similar to residential location. Generally speaking, vehicle ownership and residential location are closely related to gas price, and land use variables play a marginal role in the choice of vehicle ownership. In a research done by Cervero, the number of vehicles was regressed against socio-demographic and land use variables (Cervero 1996). This research reveals that the presence of nearby commercial land-uses is associated with relatively low vehicle ownership rates. In another model generated by Bento et al, three choices with respect to auto ownership were considered: owing zero vehicle, one vehicle, two vehicle, and three vehicles or more (Bento, Cropper et al. 2005). This study finds that households in cities with more centralized populations are more likely to own zero vehicles.

Some other papers used vehicle type as a dependent variable. Cao et al modeled north California people's choice among four vehicle types: car, minivan, SUV, and pickup (Cao, Mokhtarian et al. 2006). This research shows that traditional neighborhood designs are correlated with the choice of passenger cars, while suburban designs are associated with the choice of light duty trucks. Potoglou conducted a similar research of modeling people's vehicle type choice (Potoglou 2008). He used slightly different choice set, including: car, pickup, SUV and van. The estimates of households' latest vehicle-type choice suggest that preferences for less fuel-efficient vehicles are marginally affected by the diversity of land-uses at the place of residence, after controlling for travel to work attitudes and socio-demographic characteristics of individuals and households.

2.1.3 Mode Choice

Mode choice is the most commonly used measure of travel behavior. This is because household and zone-based mode choice proportions are easy to determine from travel surveys, and the goal of many smart growth initiatives is to reduce vehicle travel and encourage transit and non-motorized travel. In some cases (Cao et. al., 2009; Rajamani et. al., 2003; Zhang, 2006), mode choice is the only measure of travel behavior considered. In Cao et. al. (2009), three modes were considered: auto, transit, and walking or biking. Only non-work travel was examined, as non-work trips were expected to be influenced to a relatively greater degree by the built environment. Cao et. al. also employed traveler-level behavioral framework to model the relative likelihood of a traveler choosing each mode. Travelers were expected to maximize their utility, modeled as a function of travel costs, which were assumed to be fully determined by the built environment. Cao et. al. mention that this assumption, while necessary given the limited data, is a shortcoming of the study. Built environment characteristics, it is noted, may be "good predictors of non-motorized travel costs, moderate predictors for auto travel costs, but inferior predictors for transit travel costs" (Cao et. al., p. 550). Thus, socio-economic control factors may be expected to have a greater influence on transit usage than variables related to the layout of the built environment, and non-motorized mode choice modeling may yield the greatest degree of built environment sensitivity.

Rajamani et. al. (2003) considered only non-work trips in five modes: driving alone, shared ride, transit, walking, and biking. As in Cao et. al., the scope of the study is limited to non-work trips as non-work trips are flexible in destination choice, whereas only the route of a work trip may vary. Rajamani et. al. also notes that peak-period travel, typically thought of as dominated by commuters, is composed of an increasing proportion of non-work trips.

Zhang (2006) conducted a study focusing on the causes of automobile dependency, for which the sole measure of travel behavior was the probability that a traveler's only available mode was a personal vehicle. This measure may be expected to be highly correlated to mode choice, as it is presumed that, for a particular trip, a personal vehicle will be used when the only travel mode considered feasible by the traveler is a personal vehicle. Whether a personal vehicle is considered feasible is primarily a function of social and economic factors, but may also be affected by the land uses and urban form of the route of the trip, even when controlling for the length of the trip.

2.1.4 Trip Length

Trip length is another measure of travel behavior commonly used. Trip length is closely related to people's residential location and each trip's destination choice. Multiple studies indicate that higher density reduces trip length. In the paper generated by Cervero in 1996 (Cervero 1996), people's commute distance, or distance from home to workplace was regressed against socio-demographic and land use variables. People living in central city commuted around 3 fewer miles each way than those living in the suburbs. Adequate transit service also influence commute distances by influencing residential location. Vega did a study about the simultaneous choice of residential location and travel-to-work mode under central and non-central or suburban employment patterns (Vega and Reynolds-Frighan 2009). Boarnet took trip distance as one step of a two steps model in a research (Boarnet and Greenwald 2000), and the models show that neighborhoods with higher population density and higher single and multi-family attached dwelling densities have shorter trip distances, again an intuitive relationship suggesting that density reduces trip lengths.

Intra-zonal trip (trips within TAZs) proportion was used as an indirect measure of trip length by Greenwald (2006), whose goal was to improve the modeling of intrazonal travel in trip distribution models. Greenwald concluded that a traveler's decision to internalize a trip (select a destination within the origin TAZ) was significantly influenced by land use parameters of the TAZ, an influence often omitted from trip distribution in the traditional four-step model.

2.2 Characterizing the Built Environment and Smart Growth Strategies

Smart growth design principles, as articulated by the U.S. Environmental Protection Agency, include: (1) Mix land-uses; (2) Take advantage of compact building design; (3) Create a range of housing opportunities and choices; (4) Create walkable neighborhoods; (5) Foster distinctive, attractive communities with a strong sense of place; (6) Preserve open space, farmland, natural beauty, and critical environmental areas; (7) Strengthen and direct development towards existing communities; (8) Provide a variety of transportation choices; (9) Make development decisions predictable, fair, and cost effective; and (10) Encourage community and stakeholder collaboration in development decisions. (US_EPA). All these strategies can be viewed as falling into one of what are commonly referred to as the four Ds: Density, Diversity, Design and Destination. Most research trying to find the connection between smart growth strategies and travel behavior used one or combination of these "D" variables to quantify smart growth strategies. The four Ds are typically measured at the neighborhood level. Also some papers tried to use city-level variables to compare the travel behavior difference among cities. One example of such variables is city shape, which measures how much a city deviates from a circular city. (Bento, Cropper et al. 2005). A review of the land use variables used in previous research is presented next. Variables are categorized as follows: (1) Density variables; (2) Diversity variables; (3) Design variables; (4) Destination variables; and (5) city-level variables.

2.2.1 Density

Zonal measures of development *density*, typically population and employment density, are common (Zhang, 2006; Cervero and Kockelman, 1997; DKS Associates, 2007;Bento et. al., 2005) as it is expected that trips in high-density areas are shorter than trips in suburban or rural areas. Cao et. al. (2009) employed direct measures of commerce density for each neighborhood; the number of business within a certain network distance of each neighborhood were computed, as well as the distance from each household to the nearest of several types of trip attractors, such as post offices, libraries, and theaters.

Cervero and Kockleman (1997) used three measures of density: population density, employment density, and accessibility to jobs. Population density and employment density are formulated simply as the population and employment per unit area for each TAZ. Accessibility to jobs acts as a more refined measure of employment density that takes into account jobs in many nearby TAZs, rather than only a single TAZ (as is done for employment density). Each TAZ is assigned an accessibility index, defined in gravity model form:

$$Accessibility\ Index = \sum_j (jobs)_j(\exp(\lambda t_{ij}))$$

Above, i and j are the 'home' (origin) zone and destination zone, respectively, while t is the travel time and λ is the gravity model impedance factor. As interzonal commutes are common, the accessibility index can be expected to more accurately measure the employment density of a zone. This measure was later generalized into the fourth D, Destinations, of the 4-D method, and no longer used as a measure of density but rather of zonal accessibility to all travelers, rather than only commuters.

Bento et. al. (2005), in addition to population density, employed several unconventional measures of density. Bento et. al. conducted a study of households in 114 urban areas, rather than analyzing a single city, necessitating the inclusion of variables to control for characteristics of the city itself. Such variables are not present in studies of individual urban areas, but may still yield insight into the effect of large-scale urban design on travel behavior. One city-level measure of density introduced is population centrality (this will be further discussed in section 2.2.5).

Population or employment density is a common measure of the built environment in literature exploring the connection between urban form and travel behavior. Density related variables are very easy to calculate, and proven to be significant in a lot of models (Frank and Pivo 1994; Cervero and Kockelman 1997; Boarnet and Greenwald 2000; Zhang 2006; Kockelman 1997). Density's statistical significance in many models may be entirely due to its strengths as proxy variable for many difficulty-to-measure factors that influence travel behavior. For example, higher density is associated with higher access to opportunity sites, higher parking costs, higher level of congestion, and maybe higher transit access. Table 2.1 below summarizes density related variables which have been used in previous research.

Table 2.1 Review of Density Variables Used in Previous Research

Name	Definition	Reference
Gross population density	POP/AREA	(Frank and Pivo 1994; Cervero and Kockelman 1997; Boarnet and Greenwald 2000; Zhang 2006; Kockelman 1997)
Gross employment density	EMP/AREA	(Frank and Pivo 1994; Cervero and Kockelman 1997; Boarnet and Greenwald 2000; Zhang 2006; Kockelman 1997)
Gross density	(POP+EMP/AREA	(Cervero 2002)
Mixed density index	$$MDI_i = \frac{ED_i * RD_i}{ED_i + RD_i}$$ MDI_i is the mixed density index of TAZ_i. ED_i and RD_i are the employment and residential densities within a TAZ_i.	(Potoglou 2008)
Retail employment density	(Number of retail jobs) /(AREA)	(Boarnet and Greenwald 2000)
Overall neighborhood density	Several binary variables indicating if the buildings within 300ft of the surveyed unit are: single-family detached housing, low-rise multi-family buildings, mid-rise multi-family buildings, or high-rise multi-family buildings.	(Cervero 1996)
Retail or service density	Several binary variables indicating if the buildings within 300ft of the surveyed housing unit are: commercial or other non-residential buildings; grocery or drug store.	(Cervero 1996)

Housing density	Proportion of some kinds of dwelling units per unit area	(Boarnet and Greenwald 2000)

2.2.2 Diversity

Measures of the *diversity* of land uses are also useful, as mixing land uses is thought to shorten commute times and encourage trip chaining. Such measures range from simple functions of population to employment ratios (Rajamani et. al., 2003; DKS Associates, 2007; Bento et. al., 2005) to more complex methods, such as zonal entropy-based methods (Greenwald, 2006) or even raster-based methods for which land uses in each cell in a city-covering grid are compared to the land uses of adjacent cells (Cervero and Kockelman, 1997). These methods are briefly reviewed below.

Jobs-Housing Balance

Jobs-housing balance expresses the ratio of the number of jobs to population, but instead of directly using the ratio, this variable compares this ratio of a single zone with the average level of the whole area (Formula 2.1). The closer a TAZ's ratio is to the regional level, the higher that TAZ's balance index is, or closer to 1. Intuitively, higher value of balance index means a TAZ has higher potential to self-fulfill its job demand, or higher proportion of intrazonal trips. Lower balance has two possible meanings: 1) the TAZ has much more jobs than population, or 2) the TAZ has very few jobs. The first case generally means a TAZ is a major trip attractor. The second case means a TAZ is a residential-dominated area.

$$BALANCE_i = 1 - [\frac{ABS(Jobs_i - a*Pop_i)}{Jobs_i + a*Pop_i}] \qquad (2.1)$$

Where

a=the countywide ratio of number of jobs to population.

Normalized Employment to Population Ratio

This index measures the amount of employment relative to population (Formula 2.2). Instead of directly using the ratio of employment to population, the ratio is normalized by the formula to be between -1 and +1. Higher value of this index in a TAZ is considered to help it attract more HBW trips.

$$Employment\ to\ Population\ Ratio_i = \frac{Jobs_i - Pop_i}{Jobs_i + Pop_i} \qquad (2.2)$$

Entropy

Entropy is a concept from statistics, used to measure the uncertainty of a distribution, and is borrowed by transportation engineers to describe the diversity of the land uses. The measure is useful in appraising the uniformity in dispersion of a certain trait across many zones. It is defined by very simple Formula 2.3 and adopted by a lot of researchers. What's more, entropy can be used to describe anything we want to measure, such as employment type and land use type.

$$\boldsymbol{\sum_k \left[\sum_j P_{jk} ln(P_{ik})\right] / ln\ (J)\} / K} \qquad (2.3)$$

Where

P_{jk}=proportion of some subcategory variable j (such as proportion of a land use type or proportion of an employment type) within a subarea k;

J=the number of subcategories;

K=the number of subareas in a zone.

There is a weakness in this entropy definition, because maximum entropy requires that each use type takes the same proportion (*1/J*). But that case does not necessarily mean the highest mixed land use. In order to address this aspect of the definitional problem, this entropy measure can be made more regionally realistic by incorporating a weighting for each category. $P_j*ln(P_j)$ will be replaced by $Wj*P_j*ln(P_j)$ in the definition.

Dissimilarity

Dissimilarity measures the degree to which land uses abutting or diagonal to each hectare are different. It was generated by Cervero and Kockelman in a paper published in 1997 (Cervero and Kockelman 1997). They found that this dissimilarity index can increase the probability of travelling by non-single-occupant vehicle mode.

Dissimilarity is taken as a supplement to land use entropy, because land use entropy is not a very good indicator of spatial inter-mixing at a finer grain (Cervero and Kockelman 1997). A TAZ could have high land use entropy but a low value of dissimilarity. For example, if the land uses in a TAZ are isolated and miles away from each other, but each of them takes equal proportion of the total area, the dissimilarity of this TAZ will be very low, but land use entropy is 1, which is the highest possible value. This kind of land use pattern is very common in real life. For example, all the commercial land use are in a big shopping mall, residents live in a big single family residential neighborhood, and the employments are all in an industrial park.

The definition of dissimilarity is shown in Formula 2.4 and Figure 2.1. As is shown in Figure 2.1, the central "C" hectare has dissimilarity index as 5/8, for having five of its eight neighbors with distinct uses. The central "R" hectare has dissimilarity index as 4/8, for having four of its eight neighbors with distinct uses. The dissimilarity index of all the hectares in a TAZ is averaged to generate the TAZ level dissimilarity index.

$$Dissimilarity\ index = [\sum_{j=1}^{K} \sum_{l=1}^{8} (X_l/8)]/K \qquad (2.4)$$

K=number of actively developed hectare grid-cells in tract;

X_l=1 if land-use category of neighboring hectare grid-cell differs from hectare grid-cell j (0 otherwise).

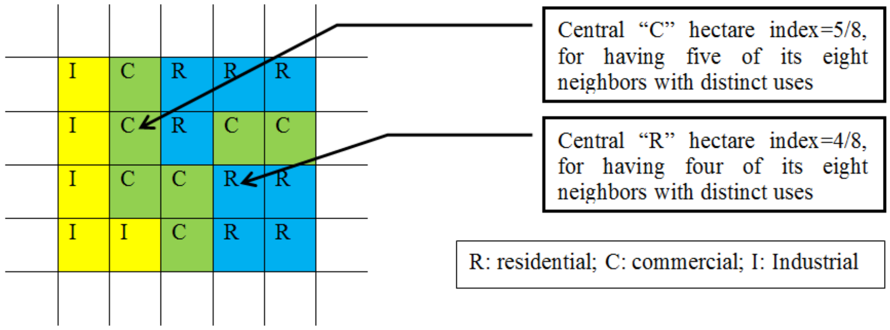


Figure 2.1 Illustration of Dissimilarity Index

The dissimilarity index in this research is based on 10 land use types: residential, retail, dinning, offices and banks, other commercial, industrial, health care, education, community and recreation.

Diversity Variables Used In Previous Research

Compared to Density, Diversity has much more and verified definitions in the previous studies. Intuitively the number of different types of business in an area could be a diversity indicator (Cao, Mokhtarian et al. 2009) ([7] in Table 2.2). Some other diversity variables measure the intensities of certain land use types, such as residential, commercial, office, industrial, and institutional, and also commercial land uses which are more finely defined: convenience stores, retail services, supermarkets, auto-oriented services, etc (Cervero and Kockelman 1997) ([4,6] in Table 2.2).

In the same paper generated by Cervero and Kockelman, vertical mixture and activity center mixture variables are included to describe other dimensions of diversity: the multiple land uses vertically in a building, and the degree of micro-mixture in an activity center (Cervero and Kockelman 1997) ([3,5] in Table 2.2). Entropy was used many times in previous research ([1] in

Table 2.2). Rajamani defined a variable very similar to entropy, called “land use mix diversity” (Rajamani, Bhat et al. 2003) ([2] in Table 2.2).

Table 2.2 Review of Diversity Variables Used in Previous Research

Name	**Definition**	**Reference**
Entropy index	Formula 2.3	(Frank and Pivo 1994)
Entropy[1]	$\sum_k \left[\sum_j P_{jk} ln(P_{ik})\right] /ln\ (J)\}/K$ P_{jk}=proportion of land-use category *j* within an area *k; J*=the number of land-use categories; and *K*=the number of subareas in a zone.	(Cervero and Kockelman 1997; Cervero 2002; Zhang 2006; Potoglou 2008; Kockelman 1997)
	Six land use categories are considered: residential, commercial, public, offices and research sites, industrial, and parks and recreation (Kockelman 1997)	
	Land use categories are denoted by employment proportions, and four of them are considered: households, retail employment, office employment, and other employment (Cervero 2002)	
	Three land use categories are considered: residential, industrial, and commercial (Zhang 2006)	
	Four land use categories within the area of walking distance (500 meters) are considered: commercial, residential, governmental, parks and industrial (Potoglou 2008)	
Land use mix diversity[2]	Equation 2.6 I2 land-use categories were used: residential, general commercial, retail and wholesale, office,	(Rajamani, Bhat et al. 2003)

	industrial, mixed commercial-industrial, health, institutional (including civic and religious), educational, ports and airports, commercial-recreational, and public parks outdoor recreational	
Dissimilarity	It gauge the degree to which uses abutting or diagonal to each hectare were different	(Cervero and Kockelman 1997; Kockelman 1997)
Vertical mixture[3]	It is the proportion of commercial/retail parcels with more than one land-use category on the site	(Cervero and Kockelman 1997)
Certain land use intensities[4]	Per developed acre intensities of land uses classified as: residential; commercial; office; industrial; institutional; parks and recreation	(Cervero and Kockelman 1997)
Activity center mixture[5]	(1) Entropy of commercial land-use categories computed across all activity centers within a zone (2) Proportion of activity centers with more than one category of commercial-retail uses (3) Proportion of activity centers with stores classified as: convenience; auto-oriented; entertainment/recreational; office; institutional; supermarkets; service-oriented	(Cervero and Kockelman 1997)
Commercial intensities[6]	Per developed acre rates of: convenience stores; retail services; supermarkets; eateries; entertainment and recreational uses; auto-oriented services; mixed parcels	(Cervero and Kockelman 1997)
Proximities to commercial-retail uses	(1) Proportion of developed acres within ¼ mile of: convenience store; retail-service use (2) Proportion of residential acres within ¼ mile of: convenience stores; retail-service use	(Cervero and Kockelman 1997)
Jobs to population balance	Retail employment and population relative to countywide ratio	(Ewing, Dumbaugh et al. 2001)

Land use mix[7]	The number of different types of businesses within specified distances.	(Cao, Mokhtarian et al. 2009)

$$\boldsymbol{Level\ of\ land\ use\ mix\ (entropy\ value) = -[single\ family \cdot log_{10}(single\ family)] + [multifamily \cdot log_{10}(multifamily)] + [retail\ and\ services \cdot log_{10}(retail\ and\ services) + [office \cdot log_{10}(office)] + [entertainment \cdot log_{10}(entertainment) + [institutional \cdot log_{10}9institutional)] + [industrial/manufacutring \cdot log_{10}(industrial/manufacutring)]} \quad (2.5)$$

$$Land\ use\ mix\ diversity = 1 - \left\{ \frac{\left|\frac{r}{T} - \frac{1}{4}\right| + \left|\frac{c}{T} - \frac{1}{4}\right| + \left|\frac{i}{T} - \frac{1}{4}\right| + \left|\frac{o}{T} - \frac{1}{4}\right|}{3/2} \right\} \quad (2.6)$$

Where r = acres in residential use (single and multi-family housing), c = acres in commercial use, i = acres in industrial use, o = acres in other land uses, and $T = r + c + i + o$. A value of 0 for this measure means that the land in the neighborhood is exclusively dedicated to a single use, while a value of 1 indicates perfect mixing of the four land uses.

Besides the above, Van Acker et. al. (2007) used a categorical, combined measure of both density and diversity. This variable is specifically design for its study area, The Netherlands, and its three central cities. Each neighborhood considered in the study is categorized along two dimensions: proximity to the triangular district formed by the three central cities, and development density (categorized as large city, regional city, small city, suburban, or rural). The resulting variable was one of several included in a function used to quantify all aspects of land use for each neighborhood. This approach to quantifying the built environment in a categorical way tailored to the study area's geography appears to be unique, and may serve to capture effects that more general measures meant to be applicable to any area may not.

2.2.3 Design

Variables that characterize street network *design* or layout are present in many studies. The primary focus of Crane and Crepeau (1998) was on the layout of streets – specifically, contrasts between grid and cul-de-sac street designs in urban areas. A one mile in diameter circle was drawn around each household for which travel survey data was available. The street network type of each circular neighborhood near the household was then subjectively determined. Researchers characterized each neighborhood's street network as grid, cul-de-sac, or mixed in design. A statistical hypothesis test found that households in cul-de-sac neighborhoods drive slightly further than households in grid neighborhoods (6.10 miles vs. 5.23 miles) with 95% confidence. Households in cul-de-sac neighborhoods also take more trips per day (2.57 trips vs. 2.08 trips), which suggests that the higher mileage may be at least partially attributable to behavioral or socio-economic characteristics. Thus, the report concluded that highly connected grid networks are not necessarily always better than cul-de-sac networks in terms of encouraging pedestrians and shortening vehicle trips.

Street density is also used (Cervero and Kockelman, 1997; DKS Associates, 2007; Bento et. al., 2005) as dense street networks are thought to encourage pedestrian travel, as well as indirectly indicate good network connectivity. Street connectivity is more directly quantified by Zhang (2006), who computed the percentage of four-way intersections in both the origin and destination TAZs of each trip. This appears to be a more methodological approach to the subjective grid/cul-de-sac categorization made by Crane and Cepeau (1998). Grid networks are made entirely from four-way intersections, while cul-de-sac networks are dominated by three-way intersections. The study found that high street network connectivity (high proportion of four-way intersections) correlates negatively with vehicle dependence, as expected.

DKS Associates (2007) used two measures related to the diversity dimension: a small-scale measure examining connectivity within TAZs (the third D, diversity, of the 4-D method) and a large-scale measure of destinations accessibility (the fourth D, destinations, of the 4-D method). The small-scale measure is meant to account for three factors affecting pedestrian travel:

$$Design = 0.0195(street\ network\ density) + 1.18(sidewalk\ completeness) + 3.63(route\ directness)$$

$$street\ network\ density = \frac{length\ of\ street}{area\ of\ neighborhood}$$

$$sidewalk\ completeness = \frac{total\ sidewalk\ centerline\ distance}{total\ street\ centerline\ distance}$$

$$route\ directness = \frac{average\ airline\ distance\ to\ center}{average\ road\ distance\ to\ center}$$

The equation appears to be empirically derived. A great deal of data and computation would be required to compute two of the three terms; sidewalk data is unlikely to be available for large areas as it is not collected by the Census Bureau as road data is, while airline and network distance computations for each road in the study area may be computationally intensive. The design variables used by previous papers are summarized in Table 2.3.

Table 2.3 Review of Design Variables in Previous Research

Name	Definition	Reference
Streets	(1) Predominant pattern (e.g. regular grid, curvilinear grid) (2) Proportion of intersections that are: four-way (proxy of grid pattern) (3) Per developed acre rates of: freeway miles within or abutting tract; number of freeway under and over-passes; number of blocks (proxy for the grain of road net); number of dead ends and cul-de-sacs (4) Averages of: arterial speed limits; street widths	(Cervero and Kockelman 1997)
Street connectivity	Percentage of four-way intersections in TAZ	(Zhang 2006)
Road density	Road length*average road width/ urbanized AREA This factor is more about the whole city, but the same concept can be applied to TAZs	(Bento, Cropper et al. 2005)
Street access	Percentage of area in 1/4 mile buffer zone covered by grid format	(Boarnet and Greenwald 2000)
Pedestrian/Bicycle facilities	(1) whether there are sidewalks in the informant's neighborhood (2) whether there are bike paths in the informant's neighborhood	(Kitamura, Mokhtarian et al. 1997)
Pedestrian and cycling provisions	(1) Proportion of blocks with: sidewalks; planting strips; street trees; overhead street lights; quadrilateral (i.e. rectangular or square) shape; bicycle lanes; mid-block crossings (2) Proportion of intersections with: signalized controls (3) Averages of: block length; sidewalk width; distance between overhead street lights; slope;	(Cervero and Kockelman 1997)

	pedestrian green lights at signalized intersections (4) Bicycle lanes per developed acre	
Pedestrian Environment Factor (PEF)	PEF score is a composite generated on four criteria: ease of street crossing, sidewalk continuity, street connectivity (grid vs. cul-de-sac) and topography. Each category is scored on a scale from one to four (four being the best ranking), so each zone has a maximum possible score of 16 and a minimum of four. Higher score means greater accommodation of non-motorized travel.	(Boarnet and Greenwald 2000)
Sidewalk provision	Ratio of sidewalk miles to road miles	(Cervero 2002)
Housing choice indicators (backyard, parking spaces available, own home)	(1) Does the informant has a private backyard or not (2) The number of parking spaces available exclusively for the informant's household use (3) whether the informant owns his/her home	(Kitamura, Mokhtarian et al. 1997)
Perceptions of neighborhood quality	Survey respondents' selection from the following reasons as why they live here: (1) No Reason to Move (2) Streets Pleasant for Walking (3) Cycling Pleasant (4) Good Local Transit Service (5) Enough Parking (6) Problems of Traffic Congestion	(Kitamura, Mokhtarian et al. 1997)
Site design	Proportion of commercial-retail and service parcels with: off-street parking; off-street parking between the store and curb; on-street front or side parking; on-site drive-ins or drive-throughs	(Cervero and Kockelman 1997)
Transit access	Whether home is within 1/2 mile of Multnomah Light Rail Corridor	(Boarnet and Greenwald 2000)

Distance to the nearest transit station		(Zhang 2006)
Transit service	The number of bus stops within walking distance	(Potoglou 2008)
Transit-oriented multi-family housing	Proportion of multi-family households in origin TAZ within one-half mile of metro-rail station	(Cervero 2002)

2.2.4 Destination

Destination related variables are also called "accessibility", representing the access of a unit to multiple opportunities, such as offices, retail service, health, and schools. The unit could be a housing unit, a building, a block, a neighborhood, or a TAZ. Some of the destination variables used in recent research are shown in Table 2.4.

Table 2.4 Review of Destination Variables in Previous Research

Destination-Related		
Name	**Definition**	**Reference**
Micro-scale accessibility indicator [1]	Distance to Nearest Bus Stop Distance to Nearest Rail Station Distance to Nearest Grocery Store Distance to Nearest Gas Station Distance to Nearest Park	(Kitamura, Mokhtarian et al. 1997)
Accessibility[6]	$$Accessibility_i = \sum_j \frac{A_j}{f(t_{ij})}$$ A_j is the attractiveness of TAZ_j, and $f(t_{ij})$ is the travel cost from TAZ_i to TAZ_j.	(Cervero and Kockelman 1997; Rajamani, Bhat et al. 2003; Kockelman 1997)
Job accessibility[2]	Number of jobs within 45-min highway network travel time (for origin TAZ)	(Cervero 2002)

Labor-force accessibility[3]	Number of households within 45-min highway network travel time (For destination TAZ)	(Cervero 2002)
Proximity[4]	Number of opportunities within walking distance. Retail and food stores, shopping-malls, schools, and entertainment facilities.	(Potoglou 2008)
Accessibility[5]	Number of establishments of each business type within specified distances and the distance to the nearest establishment of each type.	(Cao, Mokhtarian et al. 2009)

Kitamura used the distance to nearest bus stops, rail stations, grocery stores, gas stations as accessibility indicator (Kitamura, Mokhtarian et al. 1997) ([1] in Table 2.4). Many other studies calculated the number of opportunities within a certain distance from a unit as accessibility indicator (Cervero 2002; Potoglou 2008; Cao, Mokhtarian et al. 2009) ([2,3,4,5] in Table 2.4).

One of the short comings of these variables is they need a criteria, for example, the 45-min drive distance, or the walking distance. Determining such criteria needs planners' experience and not objective. This shortcoming is conquered by a more general accessibility variable, which sums the impedance-weighted number of opportunities in all TAZs (Cervero and Kockelman 1997; Rajamani, Bhat et al. 2003; Kockelman 1997) ([6] in Table 2.4).

2.2.5 City-Level Variables

Almost all the land use variables mentioned above are based on TAZs, or neighborhoods of different scales. Some research is not based local travel demand forecasting models, but compares travel patterns based on whole cities. Bento et al conducted a research based on the data from many cities of US (Bento, Cropper et al. 2005). They defined variables describing the shape of the whole city, as shown in Table 2.5.

Table 2.5 Review of City-Level Variables in Previous Research

Name	**Definition**	**Reference**
City shape	Each city is circumscribed with an ellipse equal in area to the urbanized area of the city, and the major and minor axes of the ellipse are measured. The ratio of the minor to the major axis is used the measure how much an urbanized area deviates from a circular city.	(Bento, Cropper et al. 2005)
Population centrality	It measures the distribution of population within a mono-centric city, with higher numbers indicating more centralized cities. (Equation 3.7)	(Bento, Cropper et al. 2005)
Balance of jobs versus housing	ZIP codes in each city is ordered from the one having the smallest number of jobs to the one having the largest and plot the cumulative percentage of jobs (*y*-axis) against the cumulative percentage of population (*x*-axis) to obtain a Lorenz curve. The 45-degree line represents an even distribution of jobs versus population. The balance measure is the area between the Lorenz curve and the 45-degree line, expressed as a proportion of the area under the 45-degree line. Larger values of this measure imply a less even distribution of jobs versus housing.	(Bento, Cropper et al. 2005)

$$Population\ centrality = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\sum_{i=1}^{n} P_i}{\sum_{i=1}^{N} P_i} - \frac{\sum_{i=1}^{n} P_i d_i}{\sum_{i=1}^{N} P_i d_i}\right) \quad (3.7)$$

Where

$i=1,\ldots,N$, indexes annuli around the CBD, d_i is the distance of annulus I from the CBD, and P_i is the population of annulus i.

Population centrality is a complex measure of population distribution relative to a city's central business district. Two cities with equal population densities may have different population centrality measures if one city has a dense urban core surrounded by low-density development and the other is primarily uniform, medium-density sprawl. Population centrality was found to have a significant effect on household VMT. The measure scales to the size of the city; that is, population distribution is sampled at percent distances relatives to the edge of the city, rather than at absolute distances. Where the 'edge' of a city is defined would seem to have a significant effect on the centrality measure of the city, but no formal definition of a city's edge is given in the paper. If the edge is defined subjectively, this may be a shortcoming of the population centrality measure.

2.3 Relating Travel Behavior to Smart Growth Strategies

As previously mentioned in Chapter 1, methodologies for relating travel behavior to smart growth strategies can be broadly classified into the following three categories: (1) post-processing approaches that work with a basic four-step transportation planning model; (2) modified implementations of the four-step process that can reflect various aspects of smart growth impact; and (3) disaggregate, activity-based approach. In addition, some general methods, such as correlation and regression analysis, have been used to study the impact of land-use variables on travel behavior. The aforementioned methodologies are briefly surveyed next.

2.3.1 General methods

Comparison, correlation, linear regression, and discrete choice models are among the most commonly adopted methods by researchers to find the link between travel behavior and smart growth strategies. These methods are so widely used that it would be cumbersome to list all the papers that used them. A few examples of each method are presented here.

Some early papers chose neighborhoods with different land use patterns and compared their travel behavior, and/or tested the correlation factor between land use variables and travel behavior measures. For example, Cervero compared modal splits between two distinctly different neighborhoods in the San Francisco Bay Area, and found that the neo-traditional neighborhood residents had 10% higher share of non-motorized modes than did residents of the traditional neighborhoods, for non-work trips (Cervero and Radisch 1996). For work trips, he found that compact, mixed-use and pedestrian-oriented development have strong effect on the mode of access trips to rail stations, inducing higher shares of access trips by walking and bicycle.

Correlation analysis and linear regression are both widely used methods. Kitamura et al examined the effects of land use and attitudinal characteristics on trips' mode share by building a two-step linear regression model (Kitamura, Mokhtarian et al. 1997). The first step model has socio-demographic and neighborhood characteristics as independent variables, and shows that residential density, public transit accessibility, mixed land use and the presence of sidewalk add significant explanatory power when socio-economic difference are controlled for. Then in the second model 39 attitudinal variables relating to urban life are factor-analyzed into eight factors and the eight factors are added to the models discussed above. Modeling results show that the attitudes are more strongly associated with travel behavior than land use characteristics.

Handy tested the correlations between answers to the attitudinal questions and the frequency of walking to a store (Handy 1996). The correlation factors were between 0.01 and 0.32. In another paper by Bento et al (Bento, Cropper et al. 2005), VMTs of households are used as dependent variables in some linear regression models. The study's results show that population centrality, jobs-housing balance, city shape and road density have a significant effect on annual household VMTs.

Discrete choice models are good methods in dealing with disaggregate problems. Cervero employed a binary choice to model motorized trips and non-motorized trips (Cervero 1996). Boarnet and Greenwald built ordered probit models for the trip frequency of non-work car trips (Boarnet and Greenwald 2000). Bento et al applied multinomial logit models to find the effects of urban spatial structure on mode choice among the car, bus, rail and non-motorized modes (Bento, Cropper et al. 2005). They found that probability of driving to work is lower the higher are population centrality and rail miles supplied and the lower is road density.

Cao et al employed a nested logit model for vehicle type choice among the car, minivan, SUV and pickup (Cao, Mokhtarian et al. 2006). They found that an outdoor spaciousness measure (based on perceptions of yard sizes and off-street parking availability) and commute distance impact vehicle type choice after the socio-demographic and attitudinal factors are controlled for, thus land use policies have some potential to reduce the choice of light duty trucks, thereby reduce emission.

2.3.2 Post-Processing Methods

The post processor method gets its name from the fact that it works directly on the final outputs of the four-step method planning methods, such as VMT. This is sometimes done for expediency, given the considerable time and cost of compiling local data and recalibrating large-scale models. One successful example of using the post-processing method involved examining the travel impacts of redeveloping the Atlantic Steel site in central Atlanta (Cervero 2006). The consultants used results of studies from the San Francisco Bay Area (Cervero and Kockelman 1997), metropolitan Portland, and other areas (Ewing and Cervero 2001), and found density, land use diversity, and pedestrian-friendly designs reduce vehicle trip rates and VMT.

In post-processing methods, relationships between rates of travel (VT or VMT) and descriptors of the built environment, as defined by the 4Ds previously mentioned, are investigated and the elasticities of VT or VMTs are derived. The elasticities express the percentage changes in VT and VMT as a function of percentage changes in each of the 4Ds. Two GIS-based software, INDEX and I-PLACE3S, have incorporated the 4D elasticities and have been used in evaluating the transportation benefits of alternative smart growth strategies. INDEX 4D method gives definitions for each of the 4Ds (Table 2.6), and the elasticity table (Table 2.7). The 4D elasticities are applied in a “post-processor” fashion, to a travel demand model to reflect the potential vehicle trip reduction that may result from smart-growth strategies. This has been done by application of the elasticities to aggregate measures by sub-area such as the area containing a new development, but has also been done by applying the elasticities to vehicle trip ends in a model trip table to adjust the number of trips. The revised trip table can then be used in the travel model for assignment of traffic to a roadway network to see how the trip reduction affects travel on specific links.

Table 2.6 4D Formulations of INDEX 4Ds Method

Density	*(Population + Employment)/Area*
Diversity	*1-[ABS(b*Population - Employment) / (b*Population + Employment)]* Where: *b=regional employment / regional population*
Design	*Design index=0.0195*street network density+1.18*sidewalk completeness+3.63*route directness* *Street network density*=length of street in miles/area of neighborhood in square miles *Sidewalk completeness*=total sidewalk centerline distance / total street centerline distance *Route directness*=average airline distance to center / average road distance to center
Destination	*Sum[Attractions(j)*Travel Impedance(i,j)]* for all regional TAZ *j*

Source: INDEX 4D METHOD A Quick-Response Method of Estimating Travel Impacts from Land-Use Changes Technical Memorandum

Table 2.7 4D Elasticities of INDEX 4Ds Method

	Daily VT	**Daily VMT**
Density	-0.04	-0.05
Diversity	-0.06	-0.05
Design	-0.02	-0.04
Destination	-0.03	-0.20

Source: INDEX 4D METHOD A Quick-Response Method of Estimating Travel Impacts from Land-Use Changes Technical Memorandum

2.3.3 Enhanced Travel Demand Forecasting Method

The Urban Transportation Modeling System, commonly known as the four-step method, is the primary tool used by the MPOs in U.S. for forecasting future travel demand and the performance of a transportation system. While the four-step process enjoys widespread support from decades of use, the most basic level of four-step method is developed primarily for evaluating large-scale infrastructure projects, and not for more subtle and complex smart growth strategies involving mixed land use and pedestrian friendly designs (Cervero 2006). For example, the primary unit of analysis in the four-step process, TAZs, range in size from block groups to census tracts, are too gross to gauge the fine-grained design and land-use-mix features of neighborhood-scale initiatives. To be sensitive to smart growth, measures of land uses that change under smart growth policies needed to be included in for four-step method and the comparisons of alternative programs, policies, and projects should be allowed.

A report generated by DKS Associates in 2007 provides a detailed analysis of the sensitivity of many local travel demand forecast models to smart growth strategies (DKS_Associates 2007). The limitations of the traditional four-step process, and the corresponding suggested improvements offered by DKS Associates, are shown in Table 2.8Table 2.8.

Table 2.8 Limitations of Traditional Four-step Method and Suggested Improvements

Limitation of traditional four-step method	Improvement in this research
Aggregation of zonal characteristics leads to bias in representation of trip-maker characteristics.	We use disaggregate method in trip distribution and mode split models.
Mixed-use developments are not explicitly recognized	Variables that represent mixed-use are generated and included in the models.
Land-uses are often represented by employment rather than floor area.	The land use variables are represented by both employment and parcel land use data
Transit options are inadequately represented.	A transit coverage rate is included in the method.
Non-motorized modes are not represented.	Walk and Bike are considered as two separate modes in mode choice model.
Automobile travel time is used to represent the travel cost between the OD pairs.	We include multiple measures of travel cost, including both distance and travel time of each mode.
Mode choice is only affected by time and cost characteristics.	Disaggregate mode choice model includes many socio-demographic, land use, and travel cost characteristics.

2.3.4 Activity-based Models

Activity-based models represent a significant restructuring of travel demand models. Instead of being based on trips, activity-based models structure the modeling around the activities that a household wishes to pursue during a day and how travel can be chained to satisfy the activity desires (DKS_Associates 2007). Moreover, activity-based models are disaggregate in nature, and behaviorally-based, as opposed to the statistically-based four-step process.

In an activity-based model, travel is modeled in "tours" rather than trips and the decision-making unit is the household rather than a TAZ. Activity-based modeling is an emerging method that holds promise for improving smart-growth sensitivity because it recognizes that trips made by a household are not independent of each other but are often connected for efficiency or convenience. Many smart-growth strategies are designed to reduce vehicular travel by making it easier for individuals or households to chain trips together, thus activity-based method is supposed to performs well in evaluating smart growth strategies.

Current implementations of activity-based travel demand model systems in USA include Portland, San Francisco (Jonnalagadda, Freedman et al. 2001), New Hampshire (Cambridge_Systematics_Inc 1998), Ohio, New York City, Columbus, and Atlanta(Bradley and Vovsha 2005). As is shown in Figure 2.2Figure 2.2, the activity-based model also includes four steps (DKS_Associates 2007). It uses synthetic populations based upon micro-data to do disaggregate simulations. And then the households generate complete tours or chains of trips, rather than individual trips. The trip chain generation is related to each household's socio-demographic attributes and also land use factors. The time of day of the tours is also included in the trip chain generations. After that, a joint mode/destination choice model applies.

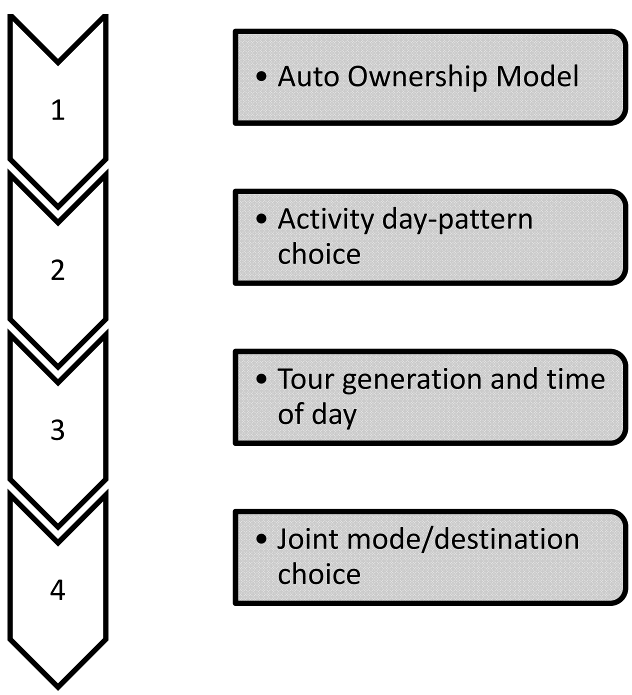


Figure 2.2 Steps of Activity-Based Model

Although activity-based model is generally considered to be more advanced method than traditional four-step method, it is not commonly applied. It is probably because building a successful activity based model needs very detailed survey data, which are often too costly to collect. While activity-based models are not investigated in great detail in this study, the study's proposed enhanced four-step planning method captures many aspects of the activity-based paradigm.

2.4 Causality in Travel Behavior Models

One reason for the mixed success of models relating the built environment to travel behavior is the phenomenon of residential self-selection; that is, residents choose to live in a place that matches their desired travel patterns, rather than adapting their patterns to match their neighborhood. For example, many studies conclude that household proximity to transit stations is well-correlated to transit usage. This may be due to people who are predisposed to transit usage choosing to live near transit stations (self-selection), or it may be that most households are willing to use transit, but only those near transit stations may do so. The distinction has consequences for planning efforts: if residential self-selection is prominent, newly constructed transit stations may see less use than expected. If residential self-selection is not prominent, it becomes easier to predict transit station use. In this way, the existence and magnitude of self-selection can greatly impact the effectiveness of travel behavior models of many types, including those that relate the built environment to travel behavior.

This phenomenon is discussed in detail by Cao et. al. (2009) who concluded that self-selection is likely influential in such modeling, and hence can create causal ambiguity. Cao et. al., in their own literature review, concluded both self-selection and land use affect pedestrian behavior to varying and difficult to quantify degrees. Additionally, Cao et. al. concluded that relatively low-density suburban neighborhoods exhibit self-selection to a greater extent than urban neighborhoods. This implies that factors such as education and income, commonly higher in suburban neighborhoods, may be correlated to the magnitude of self-selection. This matches intuition; higher income households may have more choice in where to live and which modes of transportation to use. These conclusions reinforce the earlier work of the same authors (Mokhtarian and Cao, 2008).

However, according to Naess (2009), studies that account for residential self-selection show that urban form, even when accounting for self-selection, still influences travel behavior in a significant way, and models relating the two can be statistically valid. Furthermore, Naess suggests that, when including social and economic control variables such as car ownership and mode preferences, the effects of the built environment may be underestimated. ‘Over-controlling’ is not often considered; many studies use as many control variables as may be gathered. Naess argues that control variables should be selected with caution, and in some cases, control variables such as vehicle ownership and attitudes toward transportation modes should be excluded from regression analysis, as this may lead to built environment factors being underestimated. Naess’ methodological framework consisted of five models: bivariate models relating each explanatory variable to the measure of travel behavior (daily vehicle-kilometers travelled), and four other models that controlled for various sets of factors including built environment factors, demographics, and modal attitudes. For most explanatory variables, including more controls decreased the significance of the explanatory variable, as one would expect. In other cases, the inclusion of many controls caused explanatory variables to be rendered statistically insignificant, even with a generous p-value of 0.15. In these cases, it may be that the explanatory variable was non-causal, in which case the control variables did their job well. However, Naess argues, it may also be that too many control variables were used, causing a causal and important explanatory variable to have its effects obfuscated.

Another study that specifically addresses causality problems in travel behavior models was conducted by Vance and Hedel (2007). Both a discrete choice model (mode selected for non-work trips) and a continuous choice model (distance traveled by vehicle per trip) were made, with many built environment and control variables included. Instrumental variables correlated with the built environment factors were used to aid in identifying causal relationships. An instrumental variable is a variable expected to directly affect a causally ambiguous explanatory variable. When varying an instrumental variable, a change in the dependent variable would provide evidence of a causal link between the explanatory and dependent variables. Such methods of identifying causal links are often employed when additional data is costly to obtain, such as if a new travel survey is needed. The study concluded that even with split-sample instrumental variables used, built environment factors retained their statistical significance in both the attempted models. While the study does not dismiss the effects of residential self-selection, it does state that even in the presence of these effects, built environment factors are still influential.

Greenwald (2006) applies the concept of endogeneity to travel behavior models. Endogeneity is the problem of the error term in a linear regression model being correlated to one of the regressors, which is a sign of statistical bias in the regressor's coefficient. Greenwald states that endogeneity may be a sign of residential self-selection. Instrumental variables, as used by Vance and Hedel (2007), can be used to address the problem of endogeneity.

Cervero and Kockelman (1997) discuss another source of causal ambiguity. Many of the built environment factors thought to influence travel behavior, such as mixed use development, shorter blocks, and sidewalk coverage, are correlated with one another. That is, dense urban neighborhoods will typically exhibit all, rather than some, of these traits. This multicolinearity makes it difficult to pinpoint which factors affect travel behavior, and to what extent. Linear regression model coefficients are interpreted as the estimated effect on the dependent variable of a unit change in one of the explanatory variables while the others remain constant. However, if neighborhood-level data produces two collinear variables, a statistical model cannot truly predict what would happen if one collinear variable is held constant as a control, as there is no real-world example of this.

2.5 Conclusions

This chapter has surveyed the different methods and approaches used in previous studies to quantify travel behavior and to characterize the built environment. The chapter has also surveyed methods that have recently been proposed for evaluating the impact of smart growth strategies on travel behavior, and the issue of causality in travel behavior models. The next chapters will describe how two of those proposed methods, specifically the post-processing method and the enhanced four-step planning method, were extended in the current study and implemented for the Buffalo metropolitan area. Specifically, Chapter 3 is dedicated to describing a post-processing method which was developed using travel behavior and land-use data from the Buffalo metropolitan area. Chapter 4 then describes an enhanced four-step planning method which was also developed for Buffalo, and which has the potential to increase the sensitivity of four-step panning model to the impacts of proposed smart growth strategies on travel behavior.

3. A POST-PROCESSING METHOD FOR ASSESSING THE LIKELY IMPACT OF SMART GROWTH ON TRAVEL BEHAVIOR

In this chapter, a post-processor method of quantifying and searching for relationships among many aspects of travel behavior and the built environment is developed and applied to the Buffalo, NY area. A wide scope of travel behavior is examined, and over 50 variables, many of which are based on high-detail data sources, are examined for potentially quantifying the built environment. Linear modeling is then used to relate travel behavior and the built environment, and the resulting models may be applied in a post-processor fashion to travel models to provide some measure of sensitivity to built environment modifications.

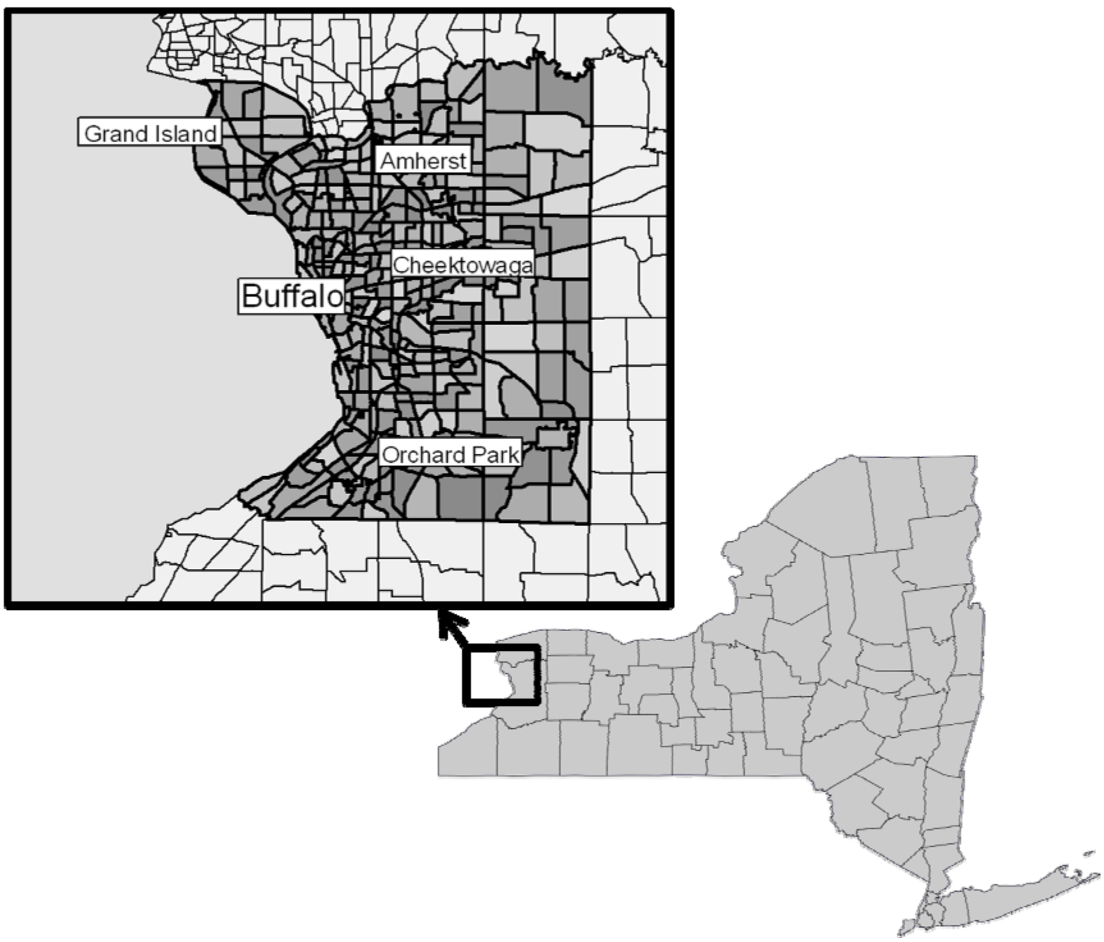
The method used to relate quantifiable spatial relationships in the built environment to aspects of travel behavior is heavily GIS-based. Spatial variability in aggregated measures of travel behavior was mapped and statistically characterized, allowing potential explanatory variables to be statistically tested for usefulness in modeling. A number of geospatial analytical concepts were used to quantify aspects of the built environment, including raster-based methods of computing point density, kernel density, and cross-tabulated polygon area. Many of the built environment variables created were parametric, allowing parameter adjustments to improve collinearity with travel behavior variables. The primary tool used for the creation and spatial analysis of each explanatory variable was ESRI ArcGIS 9; specifically, ArcMap 9.3 using an ArcInfo license and ArcEditor (ESRI, 2008).

3.1 Methodology

3.1.1 Study Area Definition

The study area used is comprised of Buffalo, NY and its neighboring communities and contains 90% of the population of Erie County, NY. There are 353 TAZs in the study area, shaded in gray in FIGURE 3.1.

FIGURE 3.1: Study area location within New York and study area TAZs



3.1.2 Quantifying Travel Behavior

For this study, zonal vehicle travel behavior was quantified in two ways: as the mean vehicle miles travelled (VMT) per zonal household, and as the mean vehicle hours travelled (VHT) per zonal household. VMT serves as a direct measure of distance travelled, while VHT accounts for variations in speed limit and congestion. Because the available data allows for both zonal VMT and VHT to be computed, both were examined for relationships to the built environment and both act as dependent variables in regression modeling. Only home-based trips were considered. Non-home-based trips were omitted from analysis due to technical constraints preventing each trip from being matched to both its origin zone built environment characteristics and home zone demographics. Thus, the scope was limited to home-based trips.

Additionally, zonal propensity for non-vehicle travel was quantified by mode choice. For this study, all modes were classified as one of three types: vehicle, transit, or non-motorized. 'Transit' includes both bus and subway trips, while 'non-motorized' includes both walking and bicycling. The zonal mode split proportions for each of these three mode types will then be computed, and used as three additional dependent variables.

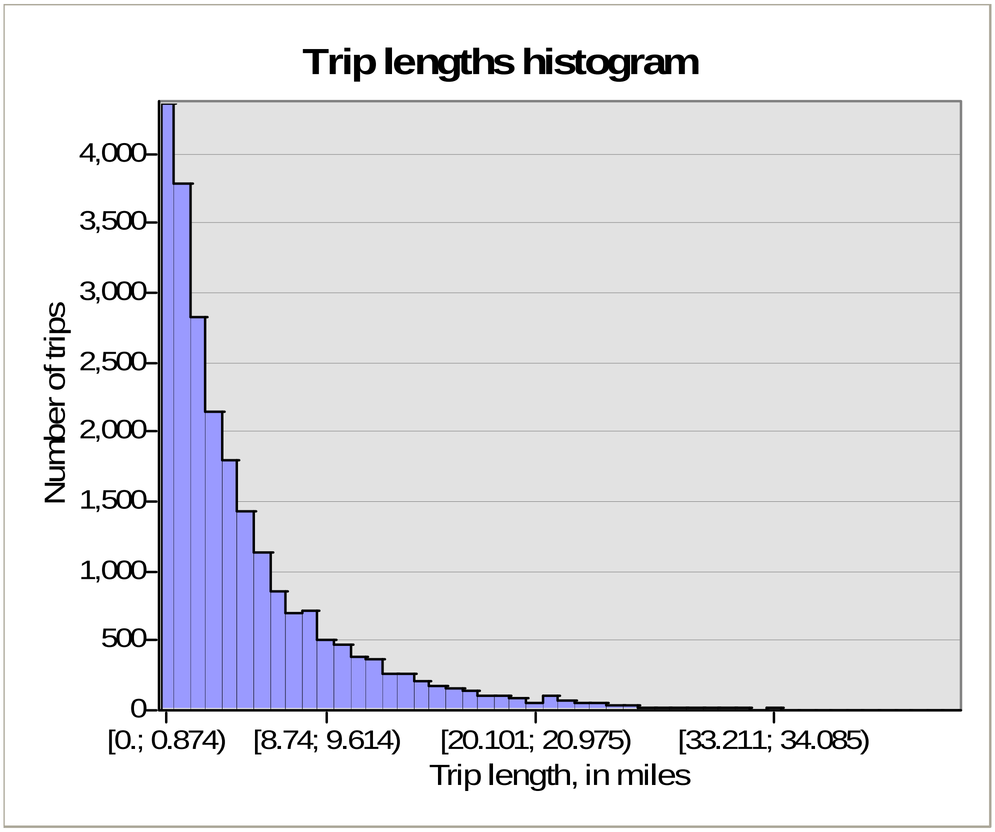
The following five variables were used to quantify different aspects of travel behavior: (1) Home-based daily VHT per household; (2) Home-based daily VMT per household; (3) Percentage of trips by vehicle; (4) Percentage of trips by transit mode; and (5) Percentage of trips by non-motorized mode.

Analysis was conducted before selecting per-household tabulations of VHT and VMT. Preliminarily, zonal total VHT and VMT were tabulated as per-resident, per-worker, per-resident and worker (the zonal sum of population and employment) and per household ratios. Two methods were used to select the best tabulation. First, a linear correlation matrix was used to determine the mean correlation between the 10 most correlated independent variables and the four sets of dependent variables. It was found that per-household measures of travel behavior correlated the best with the independent variables. Second, linear regression was used to determine the maximum possible adjusted R^2 for each tabulation method. Again, it was found that per-household dependent variables had the highest potential adjusted R^2 values, and thus the highest potential explanatory power for their linear models.

The data required to compute these dependent variables were obtained from the 2002 Regional Transportation Survey (GBNRTC, 2002). Conducted under contract to the Greater Buffalo-Niagara Regional Transportation Council (GBNRTC), the survey collected travel diaries from 2,779 homes in the Buffalo-Niagara region. Each diary tabulated trips taken by all modes for one Monday through Friday period. In total, 23,518 trips were reported. For each trip, the duration was recorded, as well as the start and end addresses. To find the distance travelled for each trip, geo-coded addresses for each start and end point were plotted in ArcMap along with a road network imported from GBNRTC's TransCAD model. An ArcMap ModelBuilder script was then used to determine the shortest network distance between each trip's start and end point, yielding the length, in miles, for each trip. These trip lengths were then used as trip VMT.

The ModelBuilder script used was taken from ESRI's Geoprocessing Resource Center, and was entitled 'MultipleRoutes' (ESRI, 2009). It requires two inputs: a road network shapefile, and a text file listing the X and Y coordinates (in the same format as the network shapefile, for example, latitude and longitude) of both the start locations and end locations of each trip. The script applies the 'Make Route Layer' tool to the network shapefile, and the 'Make XY Event Layer' tool to the start and end coordinates. Network Analyst is then used to route each trip and output a route layer. Figure 3.2, below, shows the distribution of trip lengths for all 23,518 trips recorded in the travel survey. The mean trip length was 4.7 miles.

Figure 3.2: Length of each travel survey trip, in miles



Trip durations, lengths, and modes were aggregated by household, then multiplied by expansion factors (the number of households in the zone each surveyed household represents) and aggregated by zone to obtain total daily VHT, VMT, and mode split proportions for each zone. Under-sampled zones can be expected to produce less accurate travel behavior estimates; therefore, 70 rural TAZs in Erie County with few surveyed households were excluded when determining the extent of the study area. Additionally, 39 zones from within the study area with improbable travel behavior were removed as outliers. Outlier screening was performed by first manually removing several clear outlier zones, then computing the mean per-household VHT and VMT for the remaining zones. Any zones that differed from the mean VHT or VMT by greater than three standard deviations were removed. The five travel behavior estimates were tabulated for the remaining 314 zones and used as dependent variables for the study. Figures A1 through A5 (Appendix A) display the spatial variability in zonal travel behavior, as measured by the five variables.

3.1.3 Quantifying the Built Environment

This section describes the data sources used to quantify the built environment, how the variables were defined, and the preliminary analyses performed to identify those variables that are most highly correlated to travel behavior.

Data Sources

Many characteristics of the built environment may be expected to affect travel behavior. Among those considered for the study are land uses, the street network, transit infrastructure, and the spatial distributions of population and employment. To quantify these aspects of the built environment, many data sources were used.

Land use data were gathered from a parcel map of Erie County obtained from GBNRTC. Each parcel had a three-digit land use code, the classifications for which were taken from the New York State Office of Real Property Services' Assessor's Manual (2006). Because 241 different land uses were present in the study area, categories of land uses expected to have similar influences on travel, such as 'residential' and 'commercial', were defined that included ranges of land use codes. Land use classifications, with corresponding codes, that are used in this study are listed in Table 3.1, below.

Table 3.1: Land use classifications

Land use code	Classification	Type
0	Unknown	Unknown
100-190	Agricultural	Agricultural
200	Other residential	Residential
210	Single residential	
220	Double residential	
230	Triple residential	
240-250	Rural residential	
260-283	Other residential	
300-380	Vacant	Vacant
400	Other commercial	Commercial
411	Apartments	
414-418	Living accommodations	
420-426	Dining	
430-436	Motor vehicle	
437-449	Other commercial	
450-455	Multi-use commercial	
500-546	Recreation and entertainment	Recreational
550-593	Undeveloped recreational	
600-694	Community	Community
700-744	Industrial	Industrial
800-885	Public	Public
910-972	Forest	Forest

In addition, a complete street network shapefile for Erie County from the 2000 U.S. Census was obtained, which includes minor roads typically omitted from TransCAD models and other simulation models. The inclusion of all roads is important for measurements of street density and intersection density, as is described later. The shapefile was converted into a network dataset composed of links and nodes using the 'New Network Dataset' tool in ArcCatalog.

Also obtained from GBNRTC was an employment point shapefile, showing the location and number of employees for each place of employment in Erie County. While zonal employment statistics exist, large places of employment may influence a larger area than only one zone. This effect may be captured by variables crafted from address-level employment data.

The final data source, also obtained from GBNRTC, was a transit stop point shapefile including stops for all bus and rail. This data was particularly useful in crafting variables related to mode choice. As of the time of the travel survey, 2002, four 'fare zones' were used by the Niagara Frontier Transit Authority to determine transit pricing. The fare zones were nested concentrically: fare zone 1 encompassed urban Buffalo, Fare zone 2 contained urban land surrounding fare zone 1, fare zone 3 contained suburban areas, and fare zone 4 was primarily rural.

Variable Definitions

From the parcel map, absolute measures of land use (i.e. *density* measures), such as population and employment density and the percentage of each TAZ classified as residential, commercial, and employment were calculated. The parcel map was also used to determine the parcel-covered area of each TAZ, which was used for density computations (such as population density) rather than the total area of the TAZ, which would included non-parcel-covered areas such as water. This is of particular importance for this study as many coastal TAZs extend into Lake Erie, which would skew measures of density.

For *diversity*, relative measures of land use, such as the balance of residential and commercial land for each TAZ compared to the study area residential-commercial ratio were utilized. The balance between two land use types (i.e. b_1 and b_2) is defined as:

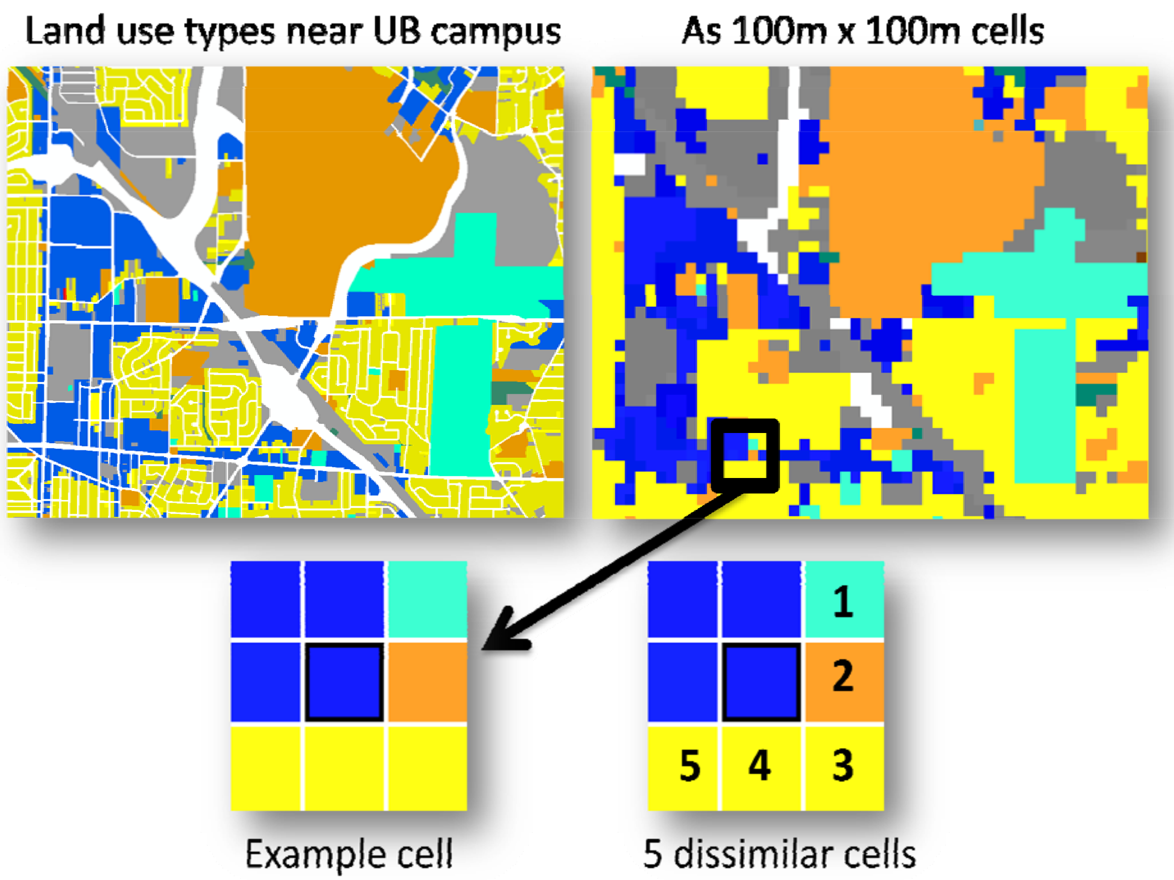
$$Balance = 1 - \left|\frac{a * b_1 - b_2}{a * b_1 + b_2}\right|$$

where $a = \frac{\sum b_2}{\sum b_1}$, the ratio of b_2 to b_1 for the entire study area

b_1 and b_2 are measures of land use, the balance of which is thought to be related to travel behavior

In addition, the study experimented with the Dissimilarity index, which was first used by Cervero and Kockelman (1997). To define that index, a 100m square grid was defined, and each cell of the grid was assigned a value corresponding to its most common land use. The land uses considered were: residential, dining, motor vehicle, offices and banks, commercial storage and distribution, other commercial, industrial, health-related, education, other community services, and recreational. All other land uses and non-parcel area were ignored. Each cell's dissimilarity value was then computed as the number of dissimilar land uses in the eight adjacent cells. High dissimilarity is a sign of land use mixing, typically considered conducive to smart growth. See Figure 3.3, below.

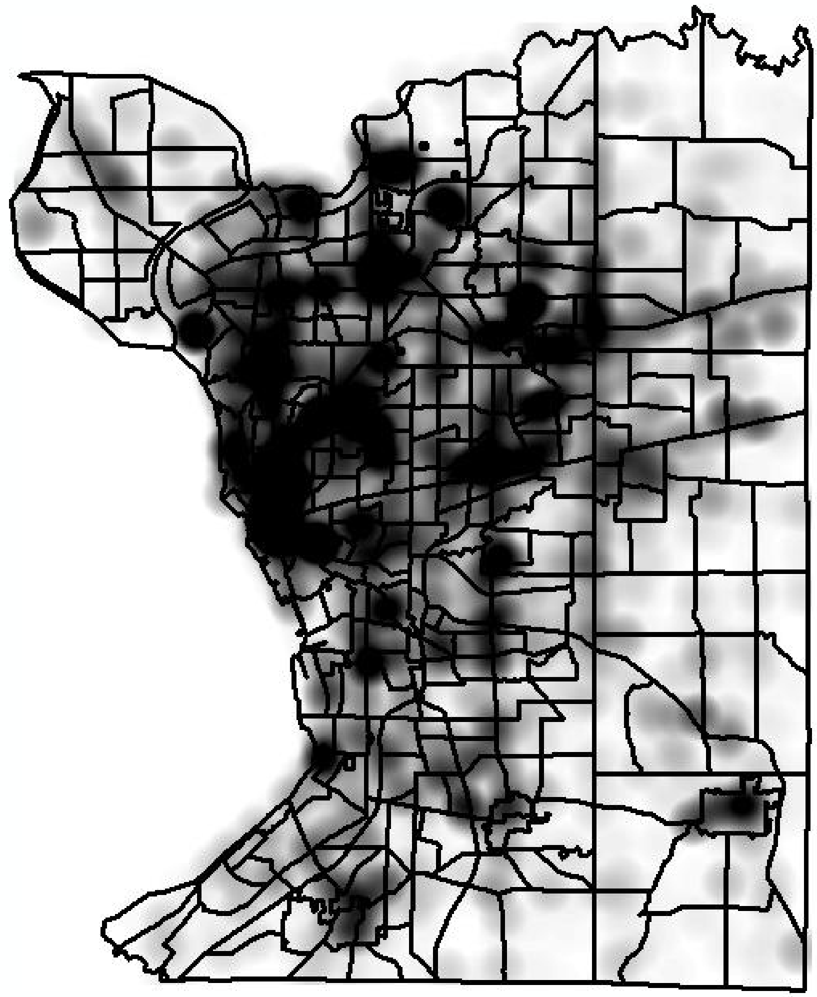
Figure 3.3: Dissimilarity index computation for a single cell



In the above figure, parcel-level land uses (upper left) are overlaid with a 100m square grid. Each cell of the grid is assigned a value corresponding to the land use with the highest proportion of the cell's area (upper right). For each hectare cell (bottom left), the eight adjacent cells are considered. The central cell is assigned a value from 0 to 8 corresponding to the number of dissimilar land uses in adjacent cells (bottom right). Cells that do not have parcel area (such as roads and water) are not considered as dissimilar to any cells. Similarly, cells whose land use with the highest area proportion is unknown, or of a type not considered, are also not considered as dissimilar to other cells.

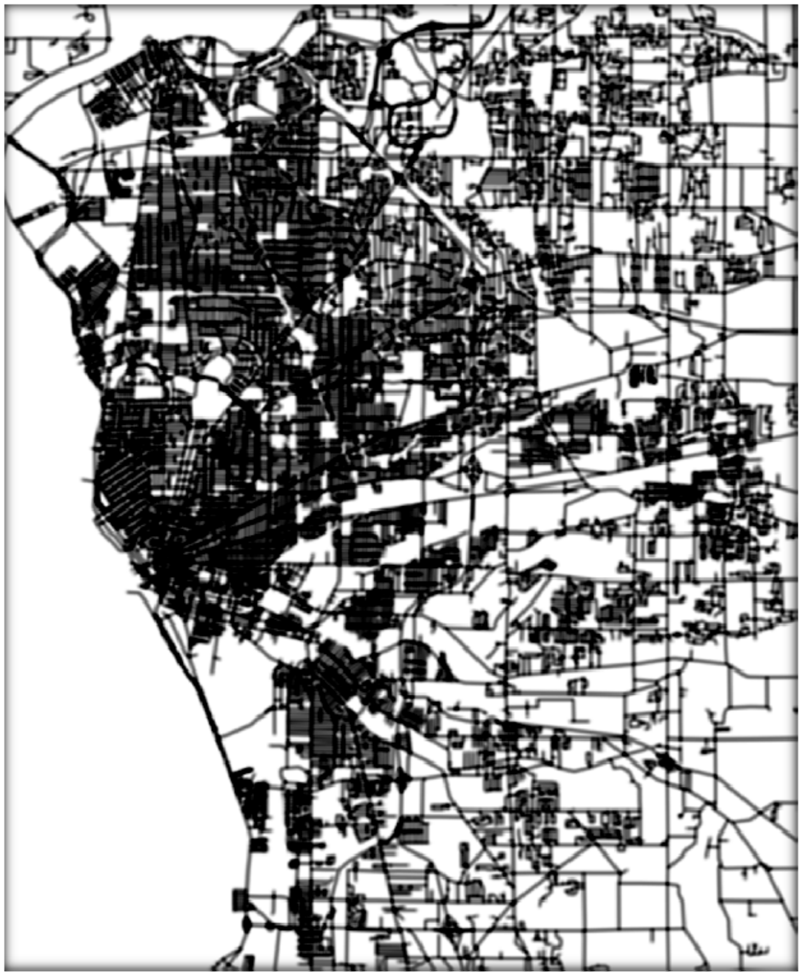
To capture aspects of network *design* and layout, variables were defined to measure design aspects such as street network density, transit stop density, and junction density. In calculating these densities (as well as other density-related variables), *spatial kernel density* was often used, as opposed to point density. For calculating the transit stops *kernel density*, for example, each transit stop is assigned a kernel radius and a kernel function whose value is highest at the source point, and decreases smoothly as the distance from the point increases until it reaches zero at a distance equal to the kernel radius. Next, in order to compute the mean kernel density of each zone, the study area was overlaid with a raster grid, and the kernel value for each cell was computed as the summation from all overlapping kernel function values. The mean kernel density for each zone was then calculated as the mean kernel value for all cells in that zone. Cells that fell on the borders of two zones were proportionally split between the zones. The advantage of using kernel density rather than point density is that the influence of built environment features that are commonly found on the boundaries of TAZs, such as transit stops or employment locations, will count, as they should, toward multiple nearby zones rather than only the zone containing the transit stop. Parameters such as kernel radius were adjusted to maximize the correlation between the kernel-based variable and travel behavior. Figure 3.4 shows an example of how the employment kernel density was calculated based on the exact employment locations and the number of employees in each.

Figure 3.4: Employment Kernel Density based on Exact Employment Locations and Number of Employees



The concept of kernel density was also applied to the layout of the street network. The street network can be modeled as a graph of links and nodes. Each link may be assigned a linear kernel function, similar to the point kernel functions used previously. The linear kernel function has its highest value at the link, and smoothly decreases to zero at the kernel radius, as pictured below:

Figure 3.5: Street network kernel density



As seen above, in urban Buffalo, where streets are parallel and close together, street kernel functions will overlap. Streets that run along the boundaries between TAZs will have a kernel function that extends into both TAZs; this may be advantageous over simpler measures of street network density for which each street only counts toward a single TAZs. In addition to street network kernel density, junction density can be computed by assigned each junction (intersection or interchange, a node) a point kernel function. Both street and junction kernel densities are aggregated zonally in the same was as employment kernel density.

Preliminary Analysis

From these data sources, over 50 variables were generated and analyzed for relationships to the travel behavior dependent variables. TAZ-level cloropleths of the dependent variables were created and examined when determining how best to quantify built environment data so the resulting variables would be well correlated to travel behavior. Explanatory variables found to have little correlation to travel behavior were discarded, while those that were correlated were iteratively improved.

Pearson (linear) and Spearman (non-linear) correlation matrices were then used as a preliminary step in variable selection. The correlation matrices listed correlation coefficients between all variables, dependent and independent, and were used to: (1) find which data sources were most promising when developing new variables; (2) diagnose the potential usefulness of independent variables for inclusion in a linear model; and (3) search for multi-collinearity among independent variables. The Spearman correlation matrix was used to search for variables that may be improved through variable transformation; however, most attempted transformations yielded marginal improvements, and the transformed variables were omitted from regression. Distribution histograms were also used; favor was given to variables with close to normal distributions. The histograms, along with scatter-plots, were also used to screen for outliers.

Candidate Variables List

Of the many explanatory variables created and tested, only 17 were eventually selected for inclusion in one of the linear models. The variables are listed in Table 3.2, and classified based upon whether they are related to *density*, *diversity* or *design* measures.

Table 3.2: Built Environment Variable List

	Variable	Brief Description
Density-related Variables	Population Density	Population divided by total parcel area
	Employment Density	Employment divided by total parcel area
	Residential proportion	% of parcel area classified as residential or apartments
	Commercial proportion	% of parcel area classified as commercial
	Employment proportion	% of parcel area classified as commercial, community or industrial, excluding apartments & community parcels that are relatively undeveloped
	High Density Index	Crafted as an indicator of urban areas, this index is the percentage of parcel area classified as apartments, two- or three-family houses, offices, retail, or multi-use. These land uses were observed to be negatively correlated to vehicle use (both vehicle mode choice and VMT).
Diversity-related Variables	Residential-Commercial Balance	The balance of residential land uses to commercial land uses
	Single-Other Residential Balance	The balance of single-family residential to all other types of residential land uses. The idea behind creating this variable is that a variety of housing options may allow employed persons to live, on average, closer to their place of employment, thus shortening commutes
	Residential-Community Balance	The balance of residential to community land uses. As community land uses include schools, health facilities, churches, and other considerable trip attractors, it is thought that mixing community services into residential areas may reduce VMT.
	Community-Commercial Balance	The balance of community and commercial land uses. Mixing these may be expected to encourage trip chaining for non-home based trips.
	Apartment-Other Residential Balance	The balance of apartments to all other types of residential land uses. As above, diverse available housing options may reduce commute times.
	Dissimilarity Index	A measure of land use diversity calculated as described previously.
	Street Network Density	The total length of street divided by the total parcel area.
	Transit Kernel	Mean kernel density for transit stops of both buses and

Design-related Variables	Density	subways.
	Junction Kernel Density	Mean kernel density for junctions (intersections and interchanges).
	Street kernel density	Mean kernel density for all streets
	Fare zone 1 Point Density	The number of fare zone 1 transit stops per unit area. The Buffalo area metro system uses multiple fare zones, with zone 1 as the most urban and zone 4 as the most rural. It was found that TAZs that were at least partially contained in zone 1 had significantly higher transit usage than those in zones 2, 3 or 4.

Household Demographics

The following five non-built-environment variables were taken from the travel survey, and are included as control variables: **Median household income, Household vehicles, Household students, Household workers,** and **Household size.** Including these variables helps address the previously-mentioned problem of residential self-selection to some extent. At the same time, however, their inclusion may result in an under-estimation of the effects of the built environment. Consider, for example, Median Household Income, which is inversely correlated to several measures of density since it is likely that high-income households self-select to relatively low density neighborhoods. Because of this, when median household income is found to be a significant explanatory variable for one of the travel behavior models, the correlated measures of density will be omitted to avoid issues with multi-collinearity. Demographic variables are included in this study for the sake of completion and comparison to previous studies.

3.2 Travel Behavior Models

To generate linear models that relate the built environment to each of the seven measures of travel behavior, ordinary least squares regression was used. Analysis was conducted in R (R Development Team, 2010), a statistical computing programming language, using the libraries ‘stats’, ‘leaps’ (Lumley, 2009) and ‘faraway’ (Faraway, 2009). The models were made exhaustive stepwise linear regression; that is, regression in which every feasible subset of regressors is attempted with a user-defined objective to either minimize or maximize. Initially, the study defined the objective as that of the maximization of the adjusted R^2, which defines the proportion of variation in the dependent variable explained by the predictors. However, it was soon found that the resulting models would be over-fit, and significant multi-collinearity would exist. After the first several steps of the stepwise regression, the models would reach over 90% of its maximum adjusted R^2, with each additional variable contributing little more to the explanatory power of the model. For some variables, the model yielding the maximum adjusted R^2 would contain as many as 22 variables. Therefore, the objective was changed to that of minimizing Mallows’ C_P, which is often used as a criterion for selecting subset regressor variables when utilizing stepwise regression (Mallows, 1973; Mallows, 1995).

3.2.1 Mallows’ C_P

Mallows’ C_P is defined as:

$$C_P = \frac{SS_{res}}{MS_{res}} - N + 2P \qquad\qquad SS_{res} = \sum_{i=1}^{N} (Y_i - Y_{pi})^2$$

Above, SS_{res} is the residual sum of squares for a model with P regressors, $MSres$ is the residual mean square when using all regressors, and N is the number of observations. Typically, it is recommended that the selected model is that with the lowest C_P for which $C_P \approx P$. However, it was noticed that some models for which $C_P < P$ could be selected with negligible loss to the adjusted R^2 value.

The models produced by this objective used significantly fewer variables, and the adjusted R^2 typically was only reduced by several hundredths of a point. Plotted below, in Figure 3.6, are the best models for non-motorized trip proportion that can be made with two or more variables. The best ten models are plotted for each number of included variables. The highest adjusted R^2 is achieved with 16 variables. However, a seven-variable model has an adjusted R^2 only 0.02 lower than the best model.

Figure 3.6: Adjusted R^2 as a function of number of included variables

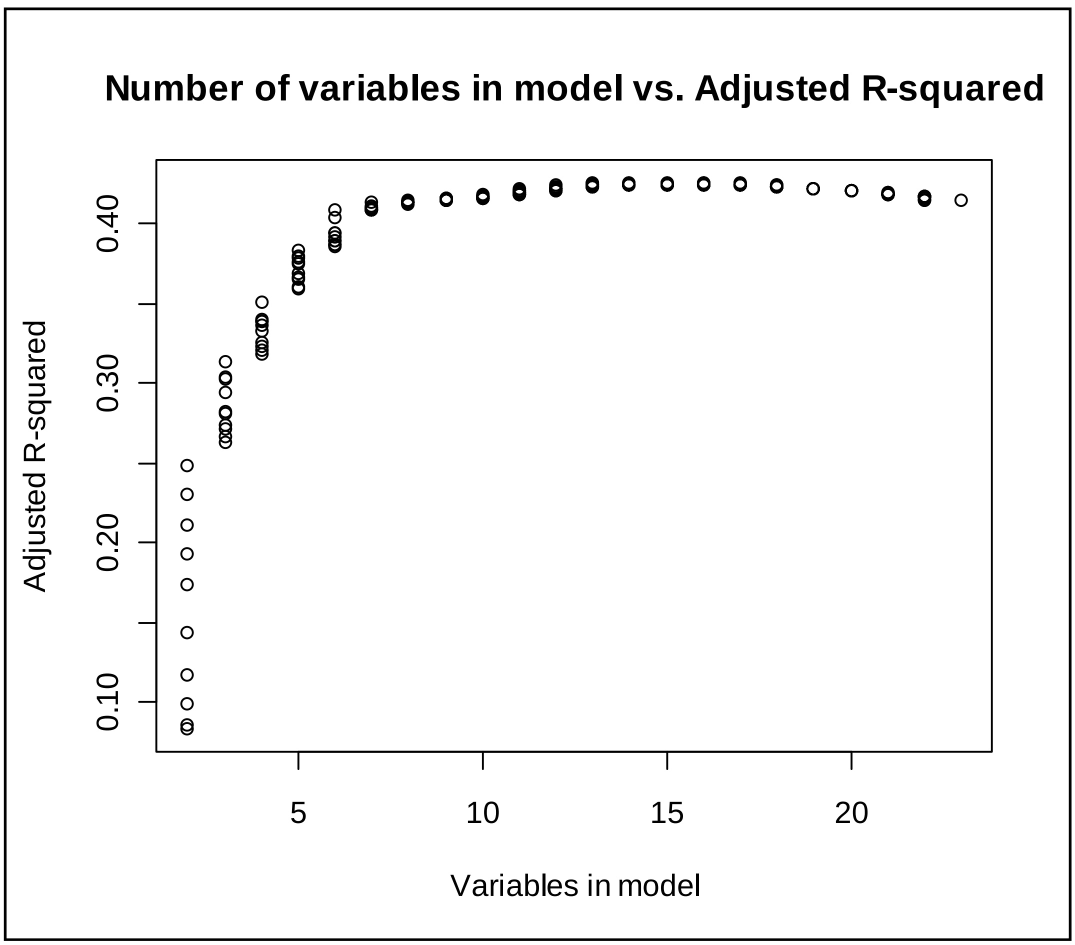
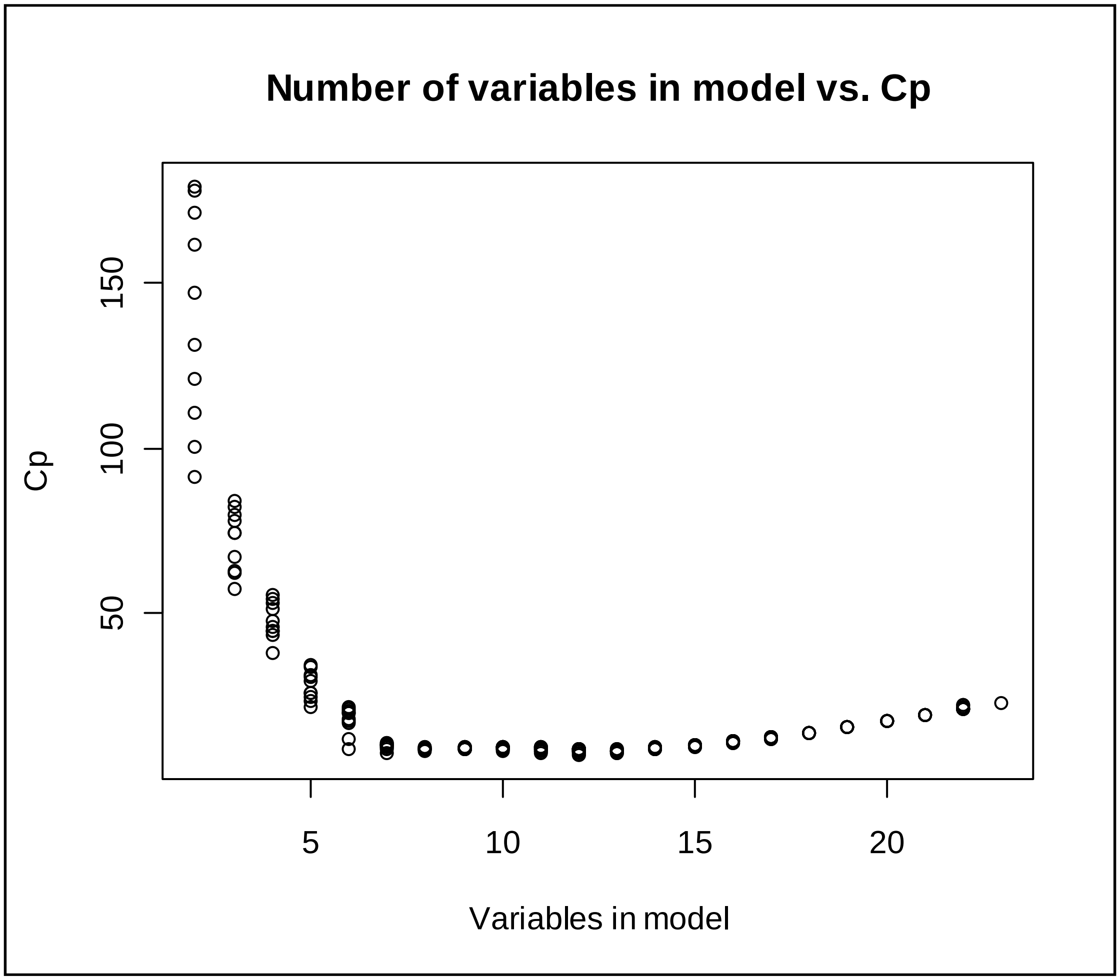


Figure 3.7: *C_P* as a function of number of included variables



Above, in Figure 3.7, are the C_P values for the same models as those plotted in Figure 3.6. The lowest C_P is achieved using a model with 11 variables and an intercept. This model has an adjusted R^2 within 0.01 of the maximum. Thus, five fewer variables can be used with negligible loss to explanatory power.

3.2.2 Regression Analysis

Tables 3.3 through 3.7 summarize the best developed models for mode choice, VHT, and VMT, respectively, and list their corresponding statistics.

Table 3.3 : Non-motorized mode choice proportion			
Variable	**Coefficient**	***β***	***t*-value**
(Intercept)	0.107		3.978
Household vehicles	-0.029	-0.195	-3.574
Household students	0.050	0.267	5.653
Population density	5.55E-06	0.300	4.015
Employment density	-1.57E-06	-0.439	-4.343
Apartment-other residential bal.	-0.060	-0.113	-2.151
Transit kernel density	1.46E-04	0.563	4.85
Junction kernel density	2.59E-04	0.154	1.779
Dissimilarity index	-0.127	-0.128	-2.031
Residential proportion	-0.091	-0.173	-2.507
Employment proportion	-0.095	-0.141	-1.975
High density index	0.207	0.217	2.817
Adjusted R^2: 0.425 Residual standard error: 0.0946 F-statistic: 22.03, p-value: < 2.2e-16			

Table 3.4: Transit mode choice proportion			
Variable	**Coefficient**	***β***	***t*-value**
(Intercept)	8.35E-03		0.402
Household vehicles	-0.031	-0.282	-3.96
Household students	-0.019	-0.136	-1.954
Household size	0.025	0.324	3.421
Single-other residential balance	-0.094	-0.155	-2.651
Apartment-other residential bal.	0.047	0.120	2.534
Street network density	1.56E-03	0.417	2.982
Transit kernel density	-7.13E-05	-0.375	-2.144
Junction kernel density	*-1.88E-04*	*-0.152*	*-1.426*
Fare zone 1 density	5.79E-03	0.575	2.672
Dissimilarity index	0.186	0.256	4.205
Commercial proportion	-0.194	-0.278	-4.033
Adjusted R^2: 0.4636 Residual standard error: 0.06693 F-statistic: 25.59, p-value: < 2.2e-16			

Table 3.5: Vehicle mode choice proportion			
Variable	**Coefficient**	***β***	***t*-value**
(Intercept)	0.657		12.895
Median household income	1.28E-06	0.129	1.99
Household vehicles	0.074	0.290	5.486
Household students	-0.054	-0.172	-3.909
Employment density	1.69E-06	0.280	3.516
Single-other residential balance	0.169	0.121	1.942
Residential-community balance	0.063	0.097	2.138
Apartment-other residential bal.	-0.114	-0.127	-2.619
Street network density	-1.51E-03	-0.176	-1.81
Fare zone 1 density	-0.010	-0.425	-3.253
Dissimilarity index	0.176	0.105	1.762
Commercial proportion	0.838	0.523	6.627
Employment proportion	-0.373	-0.328	-4.772
High density index	-0.269	-0.167	-2.755
Adjusted R^2: 0.4954 Residual standard error: 0.1493 *F*-statistic: 24.63, *p*-value: < 2.2e-16			

Table 3.6: Home-based daily VHT per household			
Variable	**Coefficient**	***β***	***t*-value**
(Intercept)	-0.061		-0.334
Median household income	7.76E-06	0.123	2.341
Household students	0.271	0.135	1.92
Household size	0.463	0.404	5.392
Community-commercial balance	*-0.307*	*-0.073*	*-1.481*
Residential proportion	0.674	0.120	2.227
Adjusted R^2: 0.3654 Residual standard error: 1.067, *F*-statistic: 37.04, *p*-value: < 2.2e-16			

Table 3.7: Home-based daily VMT per household			
Variable	**Coefficient**	**β**	***t*-value**
(Intercept)	-0.960		-0.886
Median household income	7.29E-05	0.241	4.133
Household vehicles	1.166	0.150	2.142
Household size	1.834	0.333	5.332
Street kernel density	-0.098	-0.116	-2.218
Adjusted R^2: 0.3957 Residual standard error: 5.003 F-statistic: 52.23, p-value: < 2.2e-16			

3.2.3 Discussion

As can be seen from the above, the values of the *F-statistic* for all models indicate that all developed models explain some variation in the response variable. Looking at the R^2 values, it can be concluded that the mode choice models appear to have higher explanatory value (i.e. the independent variables explain more of the variation in the response variable) compared to the VHT or VMT models. The R^2 values for the mode choice models were at least 0.425, which means that the built environment and the simple demographics included may be responsible for at least 42.5% of the zonal variation in mode choice. While the R^2 values of the developed models may still be regarded as modest or low, it should be noted that the values obtained in this study compare quite favorably to those obtained by previous studies. For example, the VMT models derived by Cervero and Kockelman (1997) had an R^2 value in the range of 0.171 to 0.203. For both VHT and VMT, the home-based models returned significantly better R^2 values than the non-home-based models. This is likely primarily due to the omission of demographics from the non-home-based models, rather than a greater influence of the built environment on home-based travel.

With respect to the significance of the individual regression coefficients or independent variables, as judged by the *t-statistic*, it can be concluded that all the independent variables are statistically significant at a 90% confidence level, with the exception of the two variables shown in italics. Moreover, most of these variables, with a few exceptions, appear to have the right sign or influence which agrees with prior intuition. Where the sign does not seem to make sense, this is probably due to the inclusion of other closely-related variables in the model that capture the same effect. For example, in the transit mode choice model, the coefficients of the street network density and fare zone 1 density variables both have a positive sign, which seems to make sense since an increase in either one of those variables should intuitively have a positive impact on transit ridership. On the other hand, the fact that the coefficient of transit kernel density variable is negative is counter-intuitive. However, it can be agreed that the former two variables (i.e. street and fare zone densities) are capturing the same effect, and hence the transit kernel density could be omitted from the model without too much loss in its explanatory power.

Focusing specifically on the built environment variables, it can be seen that built environment variables, belonging to all three categories (density, diversity, and design), appear in the mode choice models. This lends additional evidence to the hypothesis that the built environment does have an influence on mode choice, even after controlling for demographics. Fewer number of built environment variables ended up being significant in the home-based VHT or VMT models. Specifically, only one built environment-related variable was found to be statistically significant in each of the VHT and VMT home-based models, after controlling for the socio-economic variables. This suggests that the built environment has less influence on home-based travel than social and economic factors, and may be explained by the residential self-selection phenomenon. As previously mentioned, the inclusion of the median household income variable may force other density and diversity related measures out of the model because of high correlation among the variables. For the non-home-based VHT and VMT, several built environment factors are present and significant, suggesting that roughly 20% of the variation in non-home-based travel may be explained by variations in the built environment factors.

3.2.4 Elasticities

As an additional measure of the relative influence of built environment factors on travel behavior, mid-point elasticities between the built environment and travel behavior measures were computed. The use of mid-point elasticities was first proposed by Cervero and Kockelman (1997). The mid-point elasticities are computed as follows:

$$E = \beta\left(\frac{\bar{x}}{\bar{y}}\right)$$

where β is the regression coefficient, $\bar{x}$ is the mean of the built environment variable, $\bar{y}$ is the mean of the travel behavior variable.

The results are shown in Table 3.8, where elasticities of high absolute value may be expected to indicate greater influence. Variables included in linear models with relatively low significance levels were omitted from the elasticities table.

By examining the elasticities, many extrapolations regarding the influence of the built environment on travel may be made. High-density development appears to encourage non-motorized travel; however, zones of relatively high residential proportion produce fewer non-motorized trips regardless of development density. Dense street networks appear to promote transit usage. Median household income, being inversely correlated to some measures of density and land use diversity, is highly significant in the home-based VMT model, which explains the absence of many built environment variables in this model. Household vehicle ownership does not have a similar effect on the mode choice models because few built environment variables are correlated with household vehicle ownership. Zones of high dissimilarity, indicative of land use mixing, encourage transit usage to the point of creating a counterintuitive decline in non-motorized travel. This may be attributable to the fact that all three mode choice proportions must sum to 100%, and also that many survey participants may not record non-motorized trips to and from transit stops but rather record only the transit trip.

Table 3.8: Midpoint elasticities of variables present in travel behavior models

		Mode choice proportions			**Home-based daily VHT per household**	**Home-based daily VMT per household**
		Non-motorized	**Transit**	**Vehicle**		
Density	Population density	0.475				
	Employment density	-0.177		0.017		
	Residential proportion	-0.523			0.191	
	Commercial proportion		-0.647	0.124		
	Employment proportion	-0.249		-0.087		
	High density index	0.346		-0.040		
Diversity	Residential-commercial balance					
	Single-other residential balance		-1.101	0.087		
	Residential-community balance			0.038		
	Community-commercial balance				---	
	Apartment-other residential balance	-0.129	0.203	-0.022		
	Dissimilarity index	-0.499	1.482	0.062		
Design	Street network density		1.043	-0.045		
	Transit kernel density	0.209	-0.207			
	Junction kernel density	0.292	---			
	Road kernel density					-0.184

	Fare zone 1 point density		0.472	-0.035		
Demo-graphics	Median household income			0.062	0.210	0.482
	Household vehicles	-0.528	-1.133	0.118		0.252
	Household students	0.396	-0.298	-0.038	0.106	
	Household size		1.394		0.622	0.601

3.3 Principal Component Analysis

Factor analysis is a statistical analysis tool that employs factors to describe variability among explanatory variables. Factors are unobserved variables that correlate well with several explanatory variables, thus implying a previously unknown structure exists within the data. Factor analysis can be employed to reduce the number of explanatory terms needed to describe most of the variability in the data and to find relationships among explanatory variables. For the task of relating travel behavior to the built environment, both applications of factor analysis are useful: many explanatory variables may be combined into a smaller number of factors for ease of study, and explanatory variables may be found to describe variability in a common factor, thus implying they are part of a common phenomenon.

3.3.1 Factor Analysis

Precedent for the use of factor analysis comes from Cervero and Kockelman (1997), who linearly combined built environment variables into factors. It was initially conjectured that the built environment would primarily impact travel behavior along three dimensions (the three Ds – density, diversity, and design). However, it was discovered that the assumption that the three Ds were independent dimensions of travel behavior was inaccurate, and only two factors were needed to account for most of the variability among the data. The first factor, intensity, was highly correlated with density and accessibility related variables, while the second factor, walking quality, was composed of variables relating to design. Six variables were highly correlated with each factor, thus lowering the number of explanatory terms needed to explain most of the variability from 12 variables to two factors. The correlation coefficient relating each variable to its factor is referred to as the factor loading. Cervero and Kockelman obtained factor loadings of 0.796 or greater for the six variables comprising the intensity factor, indicating strong positive correlations between six measures of density and the intensity factor, and similarly strong correlations for the walking quality factor.

3.3.2 Principal Component Analysis

Related to factor analysis is principal component analysis (PCA). In PCA, each variable is considered as a dimension of the dataset. Thus, the Buffalo-area dataset used previously in linear modeling is a 22-dimensional dataset. This variable space is rotated about its many axes and a 22-dimensional unit vector is drawn, along which the dataset has its maximum possible variation. The variation in the direction of the vector is removed from the dataset, thus 'flattening' the 22-dimensional variable space into a space that, when rotated, is only 21-dimensional. Another unit vector is drawn that describes as much of the remaining variation in the data as possible, and the variation described by this data is removed. The process is repeated until 22 vectors are created. These vectors are the principal components.

Conceptually, this procedure may be thought of as a rotation of the data as a new coordinate system is chosen, with each principal component being a unit vector along one of the axes. The first principal component created would lie on the axis along which the data varies the most. Each principal component is completely uncorrelated with all of the other principal components, as they are all orthogonal to one another.

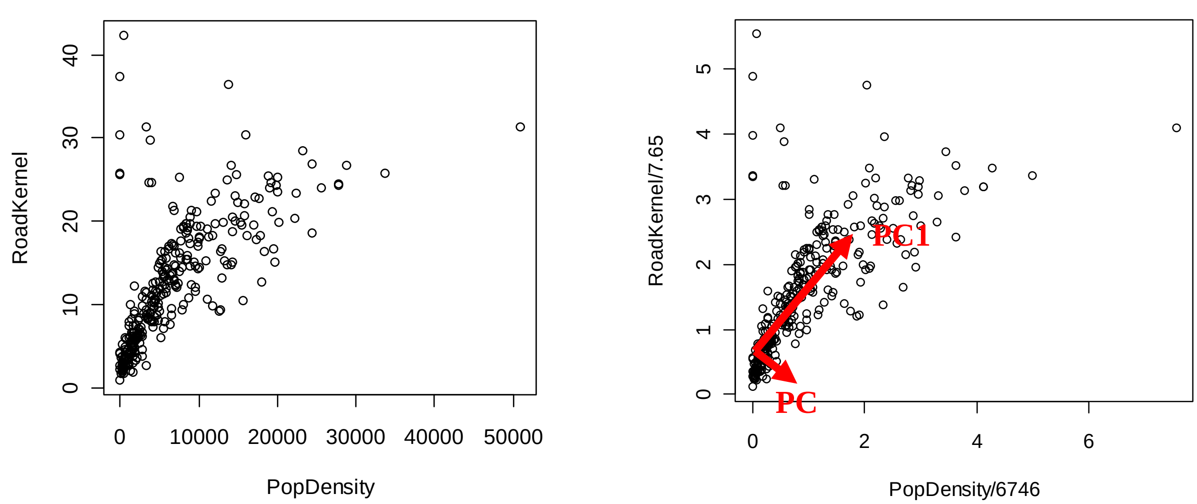
Pictured below is a two-dimensional example of PCA. Only two variables are considered: population density (the *x*-dimension) and road network kernel density (the *y*-dimension). These two variables are well-correlated, so most of the variation falls along a clear line. As the values of the variables are significantly different in magnitude, they must be scaled so as to have the same variance, one, using the scale factors in Table 3.9, below:

Table 3.9: Scale factors for example of principal component analysis

	Population Density	**Road Kernel**
Scale Factor	6746.85	7.651196

Pictured left in is a scatterplot of the two variables. Pictured right are the scaled variables and the two principal components. The first principal component describes 84.6% of the variation in the data, while the second principal component describes the remaining 15.4%.

Figure 3.8: Example of principal component analysis



Principle component analysis of all 22 variables was conducted in *R*, using the set of independent variables in Table 3.2, along with the five household demographic variables (R

Development, 2010). All variables were scaled to have unit variance, as variances differ considerably between variables that occupy different orders of magnitude. Computations were performed using singular value decomposition of the variable matrix.

Ideally, the first few principal components generated will account for almost all of the variation in the dataset, thus allowing the data to be described by a small number of principal components without much loss in accuracy. This may also imply that the data has an underlying structure in which a small number of phenomena cause most of the variation. Table 3.10 shows the proportion of the dataset variance accounted for by the first eight principal components. The first principal component can describe 36.7% of the variance of the data, and the first eight components account for 86.2% of the variance. Each additional component describes less variation, as expected.

Table 3.10: Proportion of variance attributable to each principal component

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	2.841	1.942	1.521	1.2467	1.08	0.893	0.8199	0.7863
Proportion of variance	0.367	0.171	0.105	0.0707	0.053	0.0362	0.0306	0.0281
Cumulative proportion	0.367	0.538	0.643	0.714	0.767	0.8033	0.8338	0.8619

Table 3.11, below, is the loading matrix. Each column represents the end coordinates of an origin-bound 22-dimensional unit vector, or principal component. High absolute values of the loadings indicate high correlations between the variable and the principal component.

Table 3.11: Principal component loading matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
MHHI_2000	0.256	0.001	-0.195	-0.223	0.047	-0.227	0.272	-0.002
HHVEH	0.198	-0.175	-0.358	0.053	0.164	-0.261	0.019	0.087
HWORK	0.134	-0.262	-0.361	0.212	-0.027	-0.172	-0.022	0.117
HSTUD	0.076	-0.246	-0.293	0.311	-0.208	0.384	-0.105	-0.194
HHSIZ	0.141	-0.312	-0.341	0.262	-0.069	0.12	-0.051	-0.11
PopDensity	-0.153	-0.321	0.224	-0.025	-0.306	0.153	-0.053	0.055
EmpDensity	-0.26	0.044	-0.29	-0.16	-0.035	-0.129	-0.188	0.273
ResCommBalance	0.084	-0.345	0.167	-0.027	0.189	0.065	0.058	0.56
SingleResOtherResBalance	0.25	-0.063	-0.061	-0.168	0.268	0.008	0.439	0.013
ResCommunBalance	0.052	-0.283	0.186	-0.046	0.506	0.266	-0.244	0.229
CommunCommBalance	-0.07	-0.273	0.076	-0.164	0.359	-0.333	-0.506	-0.509
ApartmentOtherResBalance	-0.073	-0.146	0.263	0.412	-0.146	-0.545	0.008	0.236
SNDbyParcelArea	-0.307	-0.13	-0.137	-0.188	-0.055	0.003	0.108	0.009
TransitKernel	-0.284	0.057	-0.298	-0.191	0.007	-0.123	-0.09	0.176
JunctionKernel	-0.29	-0.198	-0.076	-0.149	-0.033	0.126	0.085	0.032
RoadKernel	-0.259	-0.285	0.016	-0.172	-0.054	0.151	0.194	-0.062
MeanFZ1PD	-0.311	0.012	-0.225	-0.139	-0.025	-0.079	-0.061	0.114
Dissim100m	-0.237	0.005	-0.039	0.239	0.389	0.108	0.254	-0.109
ResPercent	0.127	-0.324	0.071	-0.375	-0.152	-0.118	0.282	-0.26
CommercePercent	-0.279	0.014	-0.02	0.255	0.198	-0.184	0.282	-0.147
EmpPercent	-0.254	0.139	-0.079	0.227	0.277	0.137	0.163	-0.082
HighDIndex1	-0.2	-0.271	0.219	0.161	-0.123	-0.165	0.207	-0.117

The first principal component appears to be primarily composed of variables relating to transportation infrastructure. The highest loadings belong to the variables Fare zone 1 density, street network density, junction kernel density, and transit kernel density. All of these variables are thematically similar, as they are all aspects of transportation, and all of these variables have significantly lower loadings for the second principal component. Among the lowest loadings of the first principal component are variables relating to land use balance, which are thematically different from transportation-related variables. Two commerce-related variables, employment density and commercial land area proportion, have the next two highest loadings for the first principal component. Although thematically different from the transportation variables, they are highly correlated with measures of transportation infrastructure (for example, employment density and transit kernel density are correlated with p=0.90), thus explaining their high loadings despite their lack of a thematic link. The first principal component may be tentatively assumed to represent the 'transportation infrastructure' dimension of the variable space.

The second principal component appears to be primarily composed of variables relating to population distribution. The three variables with the highest loadings are residential to commercial balance, residential land area proportion, and population density. While the transportation-related variables from the first component were highly correlated with one another, these three variables are not. For example, population density and residential land area proportion are correlated with p=0.26. This shows that structures within the data that may not be apparent from a correlation matrix can because visible through the use of PCA. Road kernel density also has a high loading here, due to the cul-de-sac nature of many suburbs that give such neighborhoods high road densities (population density and road kernel density are correlated with p=0.69). Therefore, the theme for this principal component may be tentatively assumed to be 'population density and distribution'.

Four of the five highest loadings for the third principal component belong to household demographic variables. Thus, the first three principal component may be thought to represent the following three dimensions of the built environment:

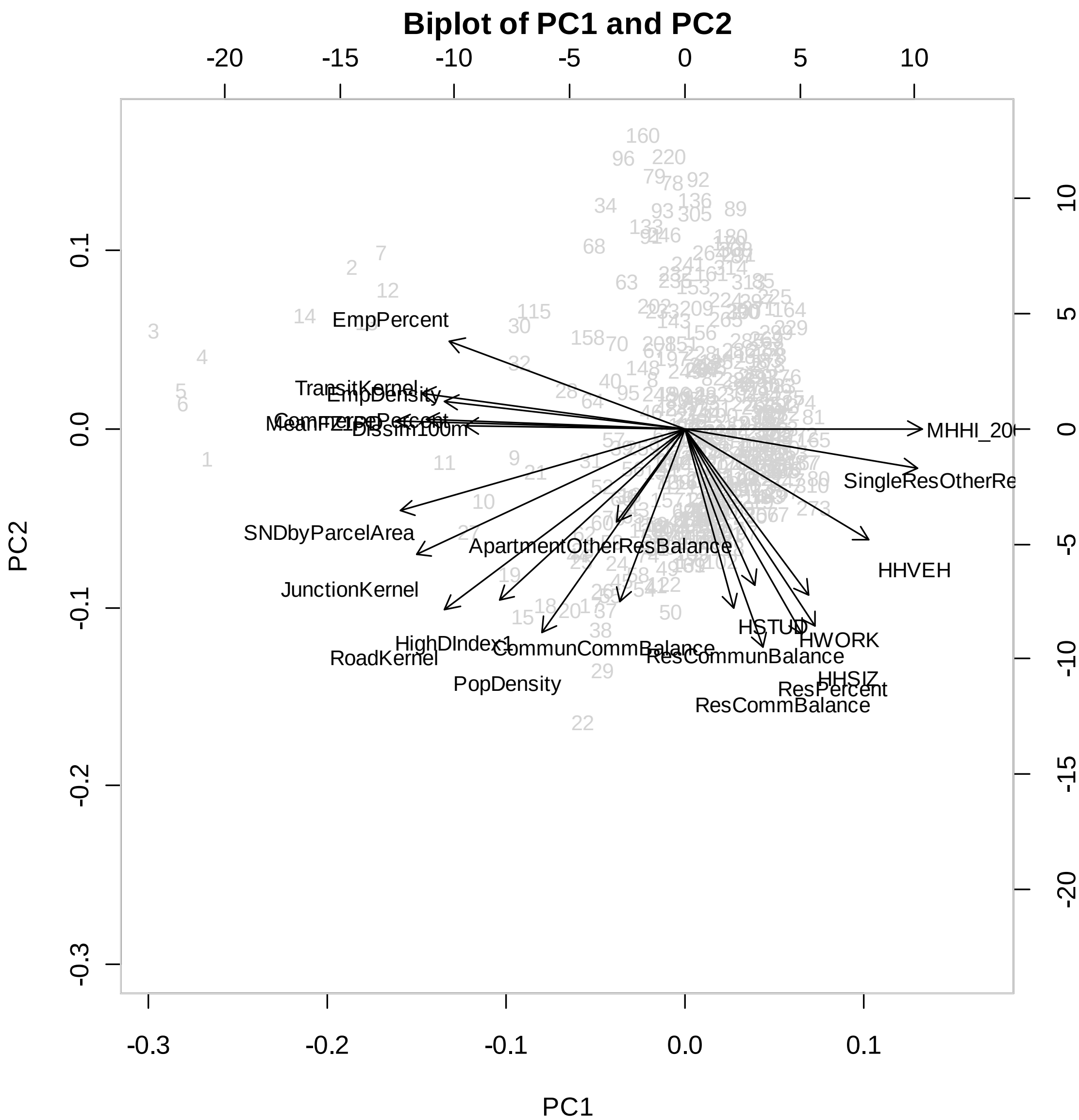
Table 3.12: Aspects characterized by first three principal components

Principal component	Built environment aspect
First	Transportation infrastructure
Second	Population density and distribution
Third	Household demographics

After the first three principal components, underlying themes can no longer be found. This is to be expected, as less variation in the data remains to be explained after the variation from each principal component is removed. These first three components account for 64.3% of the variation in the built environment data.

Another way to visualize these dimensions is through the use of principal component biplots. Below, in Figure 3.9, is a biplot of the first two principal components. A vector is drawn for each variable using the variable's loadings for principal components one and two as its end coordinates. Variable vectors pointing in either direction along the x-axis influence the first principal component, but not the second. Variables pointing along the y-axis influence the second principal component, but not the first. The magnitude of the influence is represented by the length of the vector.

Figure 3.9: Biplot of first two principal components



Principal component values for each TAZ were computed and plotted behind the vectors, in gray. The first principal component appears to account for more variation in the data due to urban TAZs (left) returning very low values for PC1. This is conceptually correct, as transportation-related variables vary more between urban and rural areas than population-related variables do (although both vary considerably).

3.4 Variable Transformation

Preliminary variable analysis conducted prior to linear regression modeling suggests that some variables may have greater explanatory power if transformed. One variable, employment kernel density, was already indirectly transformed using ArcGIS. Originally, employment kernel density spanned several orders of magnitude, as urban TAZs would be covered by many overlapping and heavily-weighted kernel functions, while rural TAZs were largely empty. The employment kernel raster was transformed with a *log* function to reduce the large spread of zonal values when aggregated. This logarithmic transformation improved the variable's correlation to the measures of travel behavior significantly.

In general, a variable is transformed by applying a continuous function to all values of the valuable. Functions are typically parametric, and parameters are chosen that maximize the transformed variable's linearity with the regressand, improve the symmetric of its distribution, and to improve the interpretability of the data prior to analysis.

3.4.1 Spearman coefficients

Evidence that other variables may be significantly improved through transformations comes from analysis of the Spearman correlation matrix (as shown in the Appendix). The matrix shows the Spearman rank correlation coefficient, ρ, between each dependent and explanatory variable. The Spearman coefficient measures how close each relation between a dependent and an independent variable is to being a monotonic function. Formally, it is the Pearson (linear) correlation of the ranked variables; that is, if the rank of each value of the variable were plotted, the Pearson coefficient of the scatterplot of the ranked variables gives the Spearman coefficient of the unranked variables. A Spearman coefficient of 1 or -1 indicates a perfectly monotonic function relating one variable to the other. The function may be increasing or decreasing, but must be purely so. A Spearman coefficient close to zero indicates that the ranked variables cannot be well related using an increasing or decreasing function.

Spearman coefficients are useful in determining if variables are good candidates for transformations. A variable may have a low Pearson coefficient, indicating a poor linear relation with a dependent variable, but a high Spearman coefficient, suggesting that a non-linear relation may exist. For example, employment density is linearly correlated with home-based VMT with a Pearson coefficient of only -0.18, but correlated with a Spearman coefficient of -0.38. Initially, two simple transformations were attempted: applying a square-root to each independent variable, and applying a base-10 *log* to each independent variable. Several variables were found to be better correlated after transformation. For example, the correlation between employment density and several travel behavior variable was improved after all values of employment density were square rooted. This is likely due to the large range of employment density; urban TAZs may employ several thousand persons per square mile, while rural TAZs may employ fewer than 10. Because of the large range and the non-normal distribution within the range, employment density may explain more variability in travel behavior when square rooted.

3.4.2 Power transformations

Rather than attempting several simple transformations such as square roots and *log* functions, a more rigorous approach was used. The approach selected was the use of a family of power transformations intended to transform all regressors closer to homoscedasticity and normal distributions, both of which are expected to improve a variable's usefulness in ordinary-least-squares regression modeling. Power-transformations are rank-preserving.

The most common type of power transformation is the Box-Cox family of parametric transformations, which take the following form:

$$y(\lambda) = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & \text{if } \lambda \neq 0; \\ \log y, & \text{if } \lambda = 0. \end{cases}$$

Above, y is the variable to be transformed and λ is the power parameter. The parameter is adjusted so as to distribute the variable symmetrically and as close to normal as possible. This form of power transformation is unable to handle negative and zero values.

3.4.3 Yeo-Johnson power transformations

Many refinements to the Box-Cox family have been proposed; perhaps the most refined is the Yeo-Johnson family, defined below:

$$y(\lambda) = \begin{cases} \dfrac{(y+1)^{\lambda} - 1}{\lambda}, & if\ \lambda \neq 0, y \geq 0; \\ \log(y+1), & if\ \lambda = 0, y \geq 0; \\ \dfrac{(1-y)^{2-\lambda} - 1}{\lambda - 2}, & if\ \lambda \neq 2, y < 0; \\ -\log(1-y), & if\ \lambda = 2, y < 0 \end{cases}$$

Above, y is the variable to be transformed and λ is the transformation parameter. The first two conditional cases are similar to the Box-Cox family, but are able to handle values of zero. The last two cases allow negative values of y to be present.

Two examples of built environment variables prior to and following the application of the Yeo-Johnson transformation are given. Population density varies considerably within the study area. Small, urban TAZs with high-rise housing have population densities in the tens of thousands of people per square mile, while suburban and rural TAZs are several orders less dense. Thus, the variable is distributed asymmetrically, and its skewed distribution means it is unlikely to be collinear with a well-distributed response variable, such as VMT per household. Figure 3.10 shows the original distribution of population density, and Figure 3.11 shows the distribution of the Yeo-Johnson transformed variable. Another example, transit kernel density, is given in Figure 3.12 (the original distribution) and Figure 3.13 (post-transformation distribution).

Figure 3.10: TAZ population density, prior to transformation

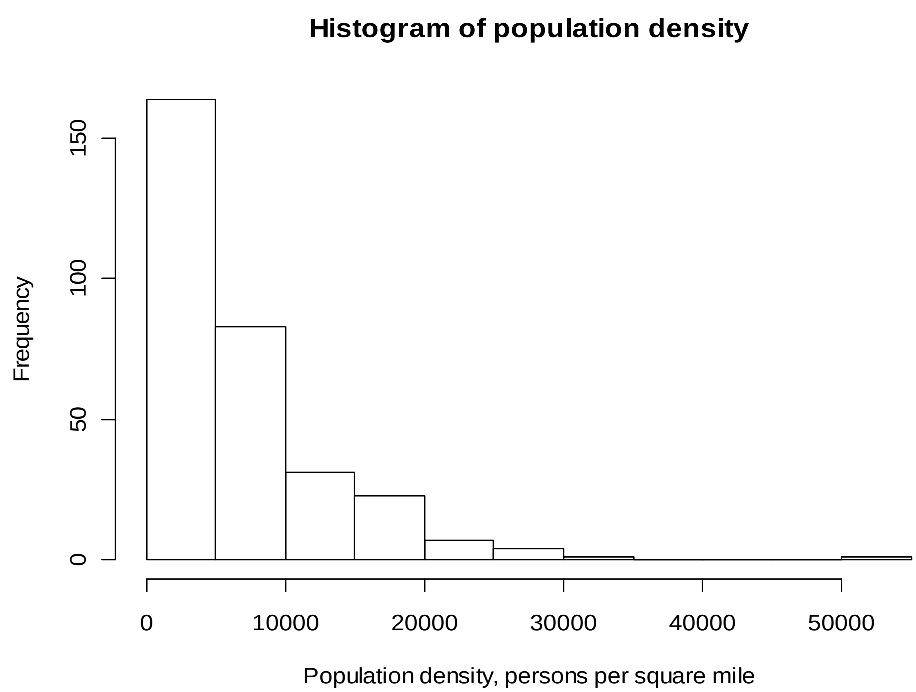


Figure 3.11: TAZ population density, following Yeo-Johnson transformation

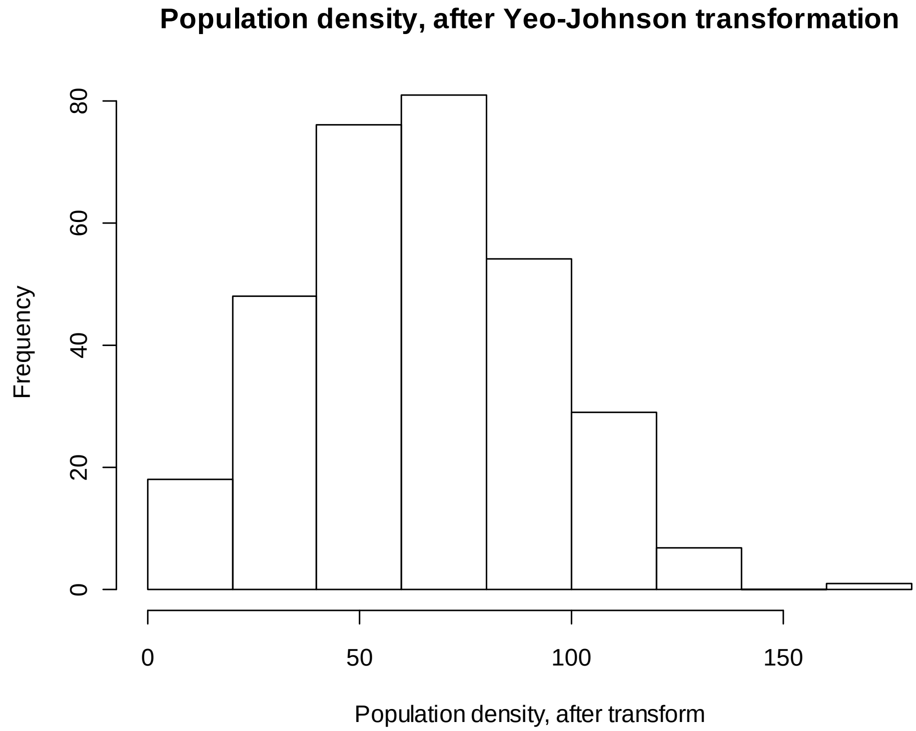


Figure 3.12: TAZ mean transit kernel density, prior to transformation

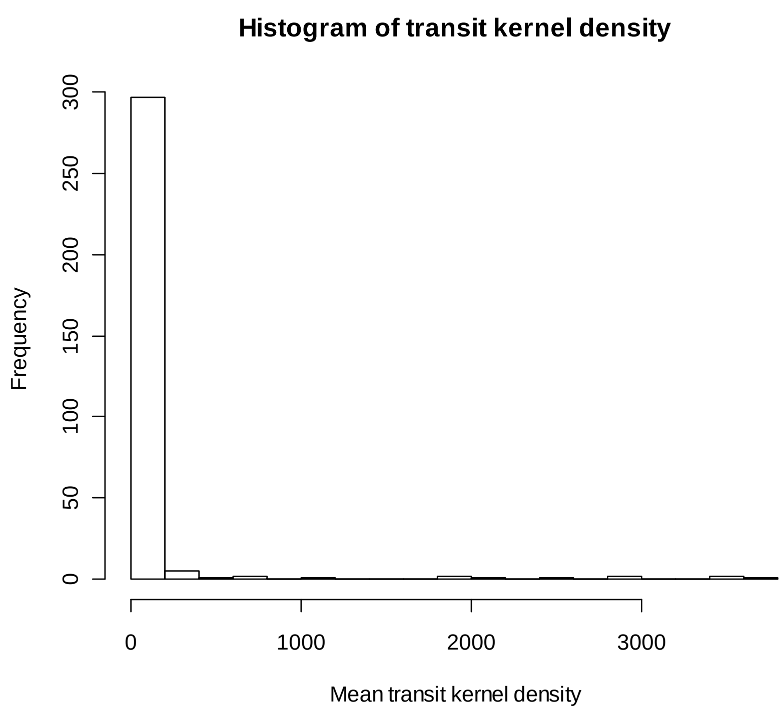
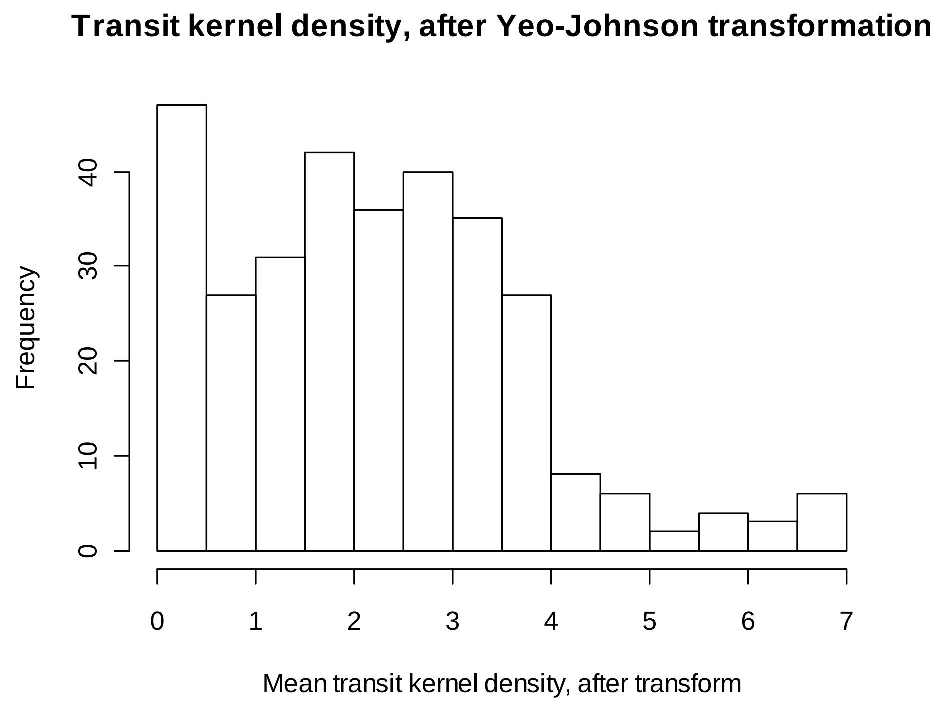


Figure 3.13: TAZ mean transit kernel density, following Yeo-Johnson transformation



As most study area TAZs have little or no transit, mean kernel density in these zones is close to zero. Downtown TAZs contain many overlapping kernel functions, resulting in mean kernel densities several orders of magnitude higher than those of suburban and rural TAZs. After the transformation is applied, the distribution is much closer to normal and likely closer to linearity with transit choice proportion, one of the dependent travel behavior variables.

Transforming to normality is not a guarantee of improved linearity with the measures of travel behavior, and it is not a guarantee that the resulting regression models will have greater explanatory power. Table 3.13, below, shows the maximum possible adjusted R^2 values that can be obtained using only the original variable space, and using only the Yeo-Johnson transformed variable space.

Table 3.13: Maximum possible adjusted R^2 of original variable space models and Yeo-Johnson transformed variable space models

	Non-transformed	Yeo-Johnson transformed
Non-motorized proportion	0.426	0.355
Transit proportion	0.465	0.429
Vehicle proportion	0.497	0.475
HB VHT per household	0.367	0.400
HB VMT per household	0.416	0.451

As seen above, explanatory power is lost in three models (the mode choice models) when all variables are transformed. Therefore, the transformation should be applied selectively, rather than to all variables. Several variables have significantly improved correlations with measures of travel behavior post-transformation, as summarized below, in Table 3.14.

Table 3.14: Correlations of selected variables with travel behavior prior to and following Yeo-Johnson transformation

Pearson correlations, prior to transformation					
	NonMSplit	TransitSplit	VehSplit	HB_VHT_pHH	HB_VMT_pHH
Household students	0.21	-0.02	-0.03	0.48	0.39
Employment density	0.21	0.38	-0.27	-0.19	-0.18
Transit kernel density	0.29	0.45	-0.35	-0.22	-0.2
Mean fare zone 1 density	0.38	0.55	-0.43	-0.23	-0.25
Pearson correlations, following Yeo-Johnson transformation					
	NonMSplit	TransitSplit	VehSplit	HB_VHT_pHH	HB_VMT_pHH
Household students	0.15	**-0.04**	0.04	**0.54**	**0.45**
Employment density	**0.31**	**0.46**	**-0.36**	**-0.28**	**-0.43**
Transit kernel density	**0.45**	**0.53**	**-0.43**	-0.2	**-0.38**
Mean fare zone 1 density	**0.52**	0.54	**-0.49**	-0.17	**-0.31**

Correlation coefficients in bold are those that have been improved by the transformation. For most variables, correlation with the five measures of travel behavior remains the same, or is improved or worsened only marginally.

To gauge the usefulness of Yeo-Johnson power transformations against simpler transformation functions, three others were attempted: *log*, square root, and the reciprocal function. For the logarithmic and reciprocal functions, a constant of 1 was added to the independent variables to prevent undefined values from being returned. For base 10 *log* transformed variables, the Yeo-Johnson transformation returned equal or better correlations for all variables except for Fare zone 1 point density, for which the base 10 *log* was slightly better. Base 2 *log* and natural (base *e*) *log* returned the same correlation matrices as base 10 *log*, indicating the base of the logarithmic function does not matter. The square root transformation was similarly outperformed by Yeo-Johnson for most variables, with the only notable exception being, again, Fare zone 1 point density. Finally, the reciprocal function was applied, and all correlation coefficients were significantly worsened, except, again, for Fare zone 1 point density (although the signs of the coefficients were reversed). It can be concluded that the Yeo-Johnson power transformation can be expected to outperform the simpler functions attempted, except in case of variables that contain a high proportion of values equal to zero. As fare zone 1 contained about 90 of the study area's 314 TAZs, roughly two-thirds of the values for this variable are zero, which likely caused poor estimation of the Yeo-Johnson power parameter, λ. For consistency, the Yeo-Johnson transformed Fare zone 1 point density will be used instead of one of the other functional forms, as the Yeo-Johnson transformed variable is still significantly better than the non-transformed original.

3.4.4 Linear regression with selective Yeo-Johnson transformations

Linear regression was re-attempted with a partially transformed variable space. The four variables in Table 6.2 (Household students, Employment density, Transit kernel density, and Fare zone 1 point density) will be Yeo-Johnson transformed, while the remaining 18 variables will remain non-transformed. As these four variables were present and highly significant in several of the original regression models, the models can be expected to be improved.

The regression method remains the same: ordinary least squares linear regression, with the objective of minimizing Mallow's C_P rather than maximizing the adjusted R^2. Analysis was again conducted in *R*, using the using the libraries 'stats', 'leaps' (Lumley, 2009) and 'faraway' (Faraway, 2009). At least one of the transformed variables was present in each of the models. The results are summarized in Table 3.15 through Table 3.19, below. Variables that have been Yeo-Johnson transformed are in bold, with their power parameter λ in parentheses. See Table B1 in Appendix B for variable definitions.

Table 3.15: Non-motorized mode choice proportion			
Variable	**Coefficient**	**β**	***t*-value**
(Intercept)	1.30E-01		3.297
HHVEH	-4.18E-02	-0.276	-3.527
HWORK	-2.37E-02	-0.131	-1.667
HHSIZ	3.95E-02	0.370	4.727
PopDensity	3.85E-06	0.208	2.835
EmpDensity (0.04)	-9.79E-03	-0.182	-2.369
ResCommBalance	3.46E-02	0.082	1.501
SNDbyParcelArea	1.82E-03	0.357	4.953
MeanFZ1PD (-1.12)	7.80E-02	0.219	3.003
Dissim100m	-1.32E-01	-0.133	-2.184
ResPercent	-7.69E-02	-0.146	-2.209
Adjusted R^2: 0.387 Residual standard error: 0.0977 F-statistic: 20.77, p-value: < 2.2e-16			

Table 3.16: Transit mode choice proportion			
Variable	**Coefficient**	**β**	***t*-value**
(Intercept)	4.41E-03		0.205
HHVEH	-2.40E-02	-0.217	-3.232
HHSIZ	1.29E-02	0.165	2.65
SingleResOtherResBalance	-1.04E-01	-0.171	-2.997
ApartmentOtherResBalance	4.64E-02	0.118	2.603
SNDbyParcelArea	1.99E-03	0.532	4.821
JunctionKernel	-2.23E-04	-0.181	-1.696
MeanFZ1PD (-1.12)	3.34E-02	0.127	2.032
Dissim100m	2.10E-01	0.289	4.891
CommercePercent	-1.88E-01	-0.269	-3.909
Adjusted R^2: 0.454 Residual standard error: 0.0674 F-statistic: 30.0, p-value: < 2.2e-16			

Table 3.17: Vehicle mode choice proportion			
Variable	**Coefficient**	***β***	***t*-value**
(Intercept)	5.93E-01		8.012
MHHI_2000	1.27E-06	0.128	1.962
HHVEH	9.70E-02	0.381	5.591
HHSIZ	-3.69E-02	-0.205	-3.361
EmpDensity (0.04)	1.76E-02	0.194	2.420
ResCommunBalance	6.16E-02	0.095	2.012
ApartmentOtherResBalance	-1.20E-01	-0.133	-2.635
SNDbyParcelArea	-3.53E-03	-0.411	-5.731
MeanFZ1PD (-1.12)	-7.05E-02	-0.117	-1.794
Dissim100m	1.81E-01	0.108	1.770
ResPercent	1.45E-01	0.164	2.297
CommercePercent	7.52E-01	0.469	5.587
EmpPercent	-3.64E-01	-0.320	-4.281
HighDIndex1	-3.32E-01	-0.207	-2.677
Adjusted R^2: 0.477 Residual standard error: 0.1519 F-statistic: 23.02, p-value: < 2.2e-16			

Table 3.18: Home-based VHT per household			
Variable	**Coefficient**	***β***	***t*-value**
(Intercept)	1.374		4.867
HSTUD (-0.64)	1.312	0.263	3.761
HHSIZ	0.389	0.339	4.787
EmpDensity (0.04)	-0.118	-0.204	-4.53
Adjusted R^2: 0.388 Residual standard error: 1.048 F-statistic: 67.17, p-value: < 2.2e-16			

Table 3.19: Home-based VMT per household			
Variable	**Coefficient**	***β***	***t*-value**
(Intercept)	5.99E+00		2.94
MHHI_2000	3.97E-05	0.130994	2.089
HHVEH	1.45E+00	0.185466	2.685
HSTUD (-0.64)	6.25E+00	0.261205	3.795
HHSIZ	7.93E-01	0.14396	1.644
EmpDensity (0.04)	-6.83E-01	-0.24672	-3.696
SNDbyParcelArea	7.42E-02	0.281957	3.067
RoadKernel	-2.26E-01	-0.2686	-2.907
Adjusted R^2: 0.4461 Residual standard error: 4.789 F-statistic: 37.01, p-value: < 2.2e-16			

3.4.5 Discussion

The adjusted R^2 values for the three mode choice models are slightly lower than those made previously with non-transformed variables (see Table 3.3 through Table 3.7). This is likely due to the highly skewed, non-normal distributions of these dependent variables (see Figures A7, A8, and A9 in Appendix A). Because the regressands are non-normal, it does not benefit the models for the regressors to be normally distributed. Thus, variables such as employment density and transit kernel density, which are positively skewed, should remain positively skewed so as to properly correspond to their effects on the response variables. This also illustrates the limited usefulness of Pearson correlation matrices in determining which variables are suited for which model. The Pearson correlation coefficients between the three mode choice variables and the four Yeo-Johnson transformed variables were improved by the transformation. However, the linear models were slightly worsened – indicating that the Pearson correlation matrix should only be used in preliminary variable screening, as it is a poor predictor of which variables will be significant in models of non-normal dependent variables.

In contrast, both the VHT and VMT models were improved by the inclusion of transformed variables. This is likely due to the relative normality of VHT and VMT distribution compared to the heavily skewed mode choice variables (see Figure A10 and Figure A11 in Appendix A). While still somewhat right-skewed, they are less skewed than explanatory variables such as transit kernel density, as seen in Figure 3.12. Thus, the transformation to normality benefits models of VHT and VMT. This is the typical assumption of linear regression – that both the dependent variable and all other independent variables are normally distributed. At least some of the missing explained variability in the models may be attributed to failures of these assumptions to hold.

To summarize, it is recommended that transformations to normality are only used when the dependent variable in the modeling is itself normal. In terms of travel modeling, this would not include variables such as mode choice, as these are likely to be skewed heavily away from normal. In urban neighborhoods with heavy transit and non-motorized travel, this may be the case. However, for a large study area, the majority of which is composed of suburban zones with little transit usage and non-motorized travel, this is not the case, and thus no transformations to normality should be used. For VHT and VMT, which have less skewed distributions, transformations to normality may be used for the entire independent variable space (see Table 3.13), or selectively, as performed above.

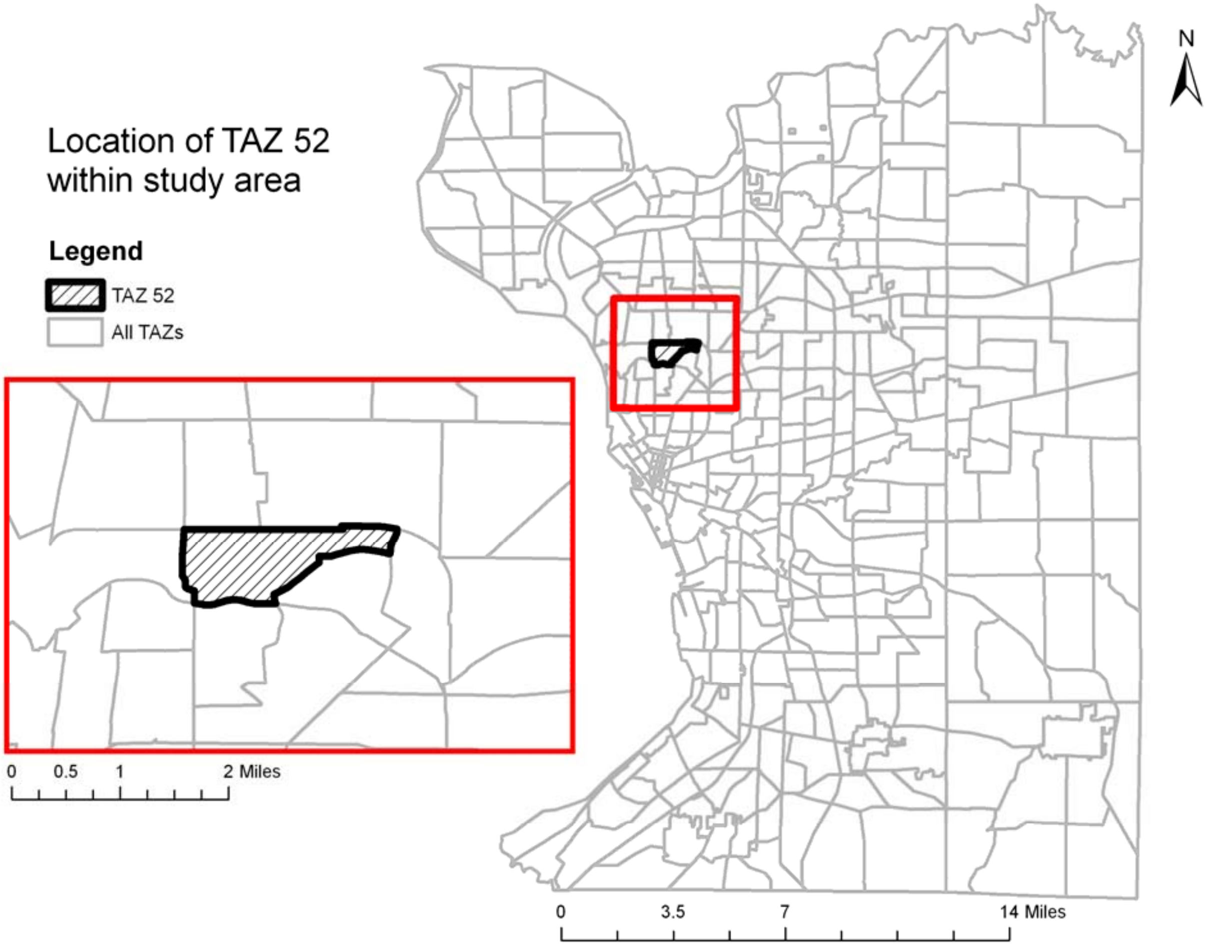
3.5 Applications to a Hypothetical Land Use Planning Scenario

In this section, a hypothetical land use and transportation planning scenario will be presented and analyzed using the models developed in section 3.2– that is, models that do not include transformed variables. The scenario developed is one typical of urban infill, in which a relatively high-density housing development is constructed on previously vacant or underutilized land. The city of Buffalo had a peak population of approximately 580,000 in 1950. As of the 2000 Census, the population had halved to approximately 293,000 (Hevesi, 2004). Population loss during this time period is a common characteristic of 'Rust Belt' cities, and is commonly attributed to loss of manufacturing jobs. Loss of population and manufacturing creates vacant land within the city. Because of this, urban infill development efforts have been undertaken in the recent years. The effects of the new housing on the network, in terms of the change in total daily VHT and VMT travelled, as well as the estimated travel mode proportions, will be estimated based on the changes in explanatory variables caused by the development.

3.5.1 TAZ 52 overview

The zone selected for analysis is TAZ 52, in northen Buffalo. The Census 2000 nation-based Traffic Analysis Zone code designation for this TAZ is 3602952. The MAF/TIGER feature class code is G6320. Its Census 2000 functional status is S, denoting it (and all other TAZs) as a statistical, rather than legal, entity (US Census Bureau, 2008). Figure 3.14 shows the location of TAZ 52 within the study area. The zone lies approximately four miles north of downtown Buffalo, and two miles east of the Niagara River, which forms the western boundary of the study area.

Figure 3.14: Location of TAZ 52 within study area



TAZ 52 is bounded to the south by the Scajaquada Expressway (NY-198) and Delaware Park, to the east by Parkside Ave., to the north by railroad tracks, and to the west by Elmwood Ave. The zone is bisected by Delaware Ave. The zone was selected for being of roughly average population and employment density for the study area, and for containing a variety of land uses. The northwest region of the zone contains commercial and industrial parcels, including storage facilities, retail, a dance hall, and a music hall. A large school occupies much of the north-central area. Much of the rest of the zone is covered by various residential parcels, with some multi-use commercial, apartments, and vacant parcels scattered throughout. See Figure C1 in Appendix C for a parcel-level land use map of the zone.

Each parcel in the entire study area is assigned a classification code by the New York State Office of Real Property Services (2006). 241 land uses were present in the study area; these can be categorized by the ten major categories (denoted by the hundreds digit of the three-digit land use code) used in the Assessors' Manual. Table 3.20 below, shows area estimates for TAZ 52. Land uses considered as 'developed' are listed in section 3.1: Methodology.

Table 3.20: Area estimates by major category for TAZ 52

Land use type	Area estimate, ft^2	% of area estimate
Unknown	1097500	7.4%
Agriculture	0	0.0%
Residential	8515000	57.7%
Vacant	347500	2.4%
Commercial	2530000	17.1%
Recreation/Entertainment	275000	1.9%
Community	1772500	12.0%
Industrial	125000	0.8%
Public	0	0.0%
Forest	92500	0.6%
Total Parcel Area	14755000	
Total Developed Area	13217500	

These areas listed above are only estimates due to the way ArcGIS calculates cross-tabulated areas. The 'Tabulate Area' tool first overlays the study area with a raster grid. Each grid cell is assigned two values: a zone value (the TAZ number), and a land use value (the three-digit land use code of the majority land use in the cell). The number of cells of each land use code is then cross-tabulated by zone. True polygonal areas of each individual parcel are infeasible on such a large scale. The smallest cell size that would not crash the computer on which ArcGIS was being run was 50 ft; therefore, each cell is 2500 ft^2, reflected in the above table. One side-effect of using raster-based area tabulations on a vector parcel map is that many of the cells contain vector detail that is lost. This is why the above table lists TAZ 52 as containing a small amount of forest area, even though no forest parcels are in the zone. An adjacent zone, containing Delaware Park, shares raster cells with TAZ 52, thus causing 0.6% of the zone to be erroneously classified as forest. This error is small, however, and the estimates may be within 1% of the true areas.

Table 3.21 and Table 3.22, below, show the zone's values for the five dependent travel behavior variables and the 22 independent variables, respectively. Descriptions of the explanatory variables may be found in Table B1 in Appendix B. The values in Table 3.21 and Table 3.22 may be compared to the summary statistics of the variables for the entire study area, in Table A1 (for travel behavior variables) and Table B2 (for explanatory variables). Zone 52 has mode choice proportions close to the study area mean, with slightly higher-than-mean VHT and VMT per household.

Table 3.21: Travel behavior variable values for TAZ 52

Travel Behavior Variable	Value
Non-motorized proportion	0.0888
Transit proportion	0.0561
Vehicle proportion	0.8550
Home-based VHT per household	1.841
Home-based VMT per household	11.088

Table 3.22: Explanatory variable values for TAZ 52

Explanatory Variable	Value
Median household income	43750
Mean household vehicles	1.6100
Mean household workers	1.2099
Mean household students	0.8175
Mean household size	2.4402
Population density	8545.0
Employment density	3887.3
Residential-commercial balance	0.8584
Single-residential to other-residential balance	0.4817
Residential-community balance	0.6893
Community-commercial balance	0.7106
Apartment to other residential balance	0.2761
Street network density	28.451
Transit kernel density	75.386
Junction kernel density	88.27
Road kernel density	15.435
Mean fare zone 1 point density	7.7963
Dissimilarity index	0.2752
Residential area proportion	0.6223
Commercial area proportion	0.1714
Employment area proportion	0.2548
High-density index	0.2112

Figure C2 in Appendix C shows the raster map used to compute the dissimilarity index of the zone. Dissimilarity values of zero cover much of the zone, as many residential cells do not border dissimilar land uses. Cells that border neighboring TAZs, as well as cells in the mix of land uses in the northwest portion of the zone, have higher dissimilarity values. The zone's dissimilarity index of 0.27 is close to the study area zonal mean of 0.30. This makes TAZ 52 particularly useful to analyze, as the baseline land use mix is close to that of the study area as a whole.

Figures C3, C4, and C5 in Appendix C map the zone's road kernel density, transit kernel density, and junction kernel density, respectively. Road kernel density is high in residential areas and low in commercial and community areas. This is typical of the entire study area, and illustrates why road kernel density is well correlated with population density (ρ =0.69). Transit kernel density varies little within the zone, but is slightly higher around the cluster of bus stops near the apartments. Junction kernel density has significant intrazonal variability in TAZ 52 and its neighboring zones due to the large gaps in the rectangular residential street network grid.

3.5.2 Scenario overview

Of particular interest are two apartment complexes in the center of the zone. These apartments were constructed in an area primarily occupied by low-density, single family residential parcels. This may be an example of 'urban' encroachment into a primarily suburban neighborhood. If the population loss of Buffalo were to stop and the city were to grow again, such developments would likely become more common as formerly suburban neighborhoods are developed to higher densities. Thus, it may be of use to determine the effects of the construction of new, high-density housing on the zone's travel behavior.

As determined in section 3.3: Principal Component Analysis, most of the variability in the explanatory variables may be thought of as occurring along three dimensions: aspects of the transportation infrastructure, aspects of population density and distribution, and household demographics (see Table 3.12). Zonal household demographics are not a built environment factor and will thus be held constant as controls for the scenario. Two land use scenarios will be developed around the first two principal components as follows:

Table 3.23: Scenario summaries

Scenario	Description	Built environment dimensions affected	Explanatory variables affected
Baseline	The current, unmodified zone	---	---
A	Low-density housing replaced with apartments	Population density and distribution (PC2)	Population density, single-residential to other-residential balance, community-commercial balance, high-density index
B	As Scenario A, also with improved transit coverage	As Scenario A (PC2), also Transportation infrastructure (PC1)	As Scenario A, also transit kernel density, mean fare zone 1 point density

Figure C6, in Appendix C, shows the hypothetical land use development plan. Two blocks near the existing apartments that are currently occupied by low-density single unit residential housing will be redeveloped into apartments. An estimated 158 residents live in these two blocks; the apartments will increase the population five-fold, to 790. This five-fold density increase estimate was obtained by estimating the ratio of the density of the existing four-floor Delaware Avenue apartments to the existing single family residences.

3.5.3 Estimation of explanatory variables

Assuming a five-fold population density increase on the two affected blocks, the explanatory variables can be re-computed for Scenario A. The zonal population density will increase by 16%, several land use balance variables will change, and the high density index will increase by 24%. For Scenario B, the same land use changes will occur, along with improvements in transit coverage that will double the zone's mean transit kernel density and fare zone 1 point density. As the existing Delaware Avenue apartments are served by clusters of transit stops (see Figure C4 in Appendix C), the improvement of transit coverage fits the proposed development of new apartments. Table 3.24, below, summarizes the changes in explanatory variables estimated for both scenarios:

Table 3.24: Estimated changes in explanatory values for hypothetical scenarios

Explanatory Variable	Baseline value	Scenario A	Scenario B
Median household income	43750	43750	43750
Mean household vehicles	1.610	1.610	1.610
Mean household workers	1.210	1.210	1.210
Mean household students	0.818	0.818	0.818
Mean household size	2.440	2.440	2.440
Population density	8545	9883	9883
Employment density	3887	3887	3887
Residential-commercial balance	0.858	0.858	0.858
Single-residential to other-residential balance	0.482	0.440	0.440
Residential-community balance	0.689	0.689	0.689
Community-commercial balance	0.711	0.711	0.711
Apartment to other residential balance	0.276	0.542	0.542
Street network density	28.451	28.451	28.451
Transit kernel density	75.38	75.38	150.77
Junction kernel density	88.27	88.27	88.27
Road kernel density	15.43	15.43	15.43
Mean fare zone 1 point density	7.796	7.796	15.59
Dissimilarity index	0.275	0.275	0.275
Residential area proportion	0.622	0.622	0.622
Commercial area proportion	0.171	0.171	0.171
Employment area proportion	0.255	0.255	0.255
High-density index	0.211	0.262	0.262

Highlighted in blue in Table 3.24 are variables related to population density and distribution that are changed in both Scenario A and Scenario B. Highlighted in red are transit-related variables affected only by Scenario B. Six of the 17 built environment variables are affected by Scenario B. The remaining 11 built environment variables are those related to employment, community land uses, and the street network that are not expected to be significantly impacted by the scenarios.

3.5.4 Estimation of travel behavior

Because the linear models for home-based VHT and VMT do not include any of the six affected explanatory variables, these two models will be omitted from the following analysis. The three mode choice variables will be estimated for the baseline scenario (existing conditions), Scenario A (only land use changes), and Scenario B (land use and transit changes). Analysis was conducted using the predict.lm function in R, yielding fitted values summarized in Table 3.25, below:

Table 3.25: Mode choice fitted values

Travel behavior variable	Actual value	Fitted value (baseline)	Fitted value (scenario A)	Fitted value (scenario B)
Non-motorized proportion	0.089	0.087	0.089	0.100
Transit proportion	0.056	0.058	0.075	0.114
Vehicle proportion	0.855	0.808	0.757	0.680

As the three mode choice linear models were estimated independently of one another, the fitted values are not guaranteed to sum to 1. Thus, a scale factor can be applied to each fitted value to enforce this constraint. The scale factor for each is computed as follows:

$$Scale\ factor = \frac{1}{\sum Non-motoized, transit, vehicle\ proportions}$$

After scaling, the following mode choice estimates for each scenario are obtained:

Table 3.26: Scaled mode choice fitted values

Travel behavior variable	Actual value	Fitted value (baseline)	Fitted value (scenario A)	Fitted value (scenario B)
Non-motorized proportion	0.088838	0.091344	0.096681	0.111796
Transit proportion	0.056162	0.061009	0.081029	0.127841
Vehicle proportion	0.855001	0.847647	0.82229	0.760362

As seen above in Table 3.26, the scaled fitted value for the baseline, existing conditions in TAZ 52 closely match those reported by the travel survey (actual values). All three mode proportions are within 1% of the actual values, thus validating the models for this zone.

The models predict that, for Scenario A, non-motorized trip proportion will remain roughly the same, while 2% of the zone's trips will be taken by transit rather than vehicles. The shift in mode choice predicted for Scenario B is more drastic; vehicle proportion will drop by 9%, transit proportion will increase by 7%, and non-motorized proportion will increase by 2%. According to the TransCAD model of Erie County, TAZ 52 generates an estimated 20,619 trips each day (about 12 per household). Table 3.27, below, shows the number of trips taken by each mode, as estimated by the travel survey and the linear models (baseline scenario), as well as Scenario A and Scenario B. Also listed are the percent changes relative to the baseline scenario.

Table 3.27: Estimated number of trips by mode

	Travel survey	Baseline scenario	Scenario A		Scenario B	
Travel behavior variable	**Estimate**	**Estimate**	**Estimate**	**% change**	**Estimate**	**% change**
Non-motorized trips	1832	1883	1993	5.8%	2305	22.4%
Transit trips	1158	1258	1671	32.8%	2636	109.5%
Vehicle trips	17629	17478	16955	-3.0%	15678	-10.3%

The predictions made in Table 3.27 are based on the assumption that neither scenario will change the total number of trips taken in the zone, but rather only the mode choice proportions. This assumption is somewhat unrealistic, and was made only to illustrate the absolute magnitude of the change in zonal travel, measured in number of trips, that may be attributable to the scenarios. Under Scenario B, for example, the number of transit trips would double, while 10% fewer vehicle trips would be made. Even when the transit infrastructure is not altered, as in Scenario A, 33% more transit trips are taken, and a modest 3% reduction in vehicle trips can be seen. This is attributable to apartment dwellers being less likely to own vehicles and more likely to use transit. This, in turn, may be due to transit coverage being maximized by placing transit stops in densely populated regions, such as between the apartment complexes in TAZ 52 (see Figure C4 in Appendix C). This may also be due, in part, to the relatively low income of apartment dwellers.

3.6 Summary

In this chapter, a post-processor method of quantifying the impact of the built environment and smart growth strategies on travel behavior was developed. The models developed were then applied to a hypothetical land use and transportation planning scenario. The next chapter will describe the second approach pursued in this study in order to develop planning tools that are sensitive to the built environment and land-use variables, namely an enhanced four-step planning process.

4. ENHANCED TRAVEL DEMAND FORECASTNG PROCESS SENSITIVE TO SMART GROWTH STRATEGIES

This chapter will describe an enhanced four step, travel demand forecasting process which was developed to allow the process to better model and reflect the impact of proposed smart growth strategies on travel behavior. This chapter begins with a description of the variables used in this part of the study to quantify land use. This is followed by the enhanced models developed by the study for the trip generation, trip distribution and mode choice steps, respectively. Finally, a case study will be presented to demonstrate the feasibility of applying the enhanced travel demand forecasting method to assess the impact of different smart growth land use scenarios.

4.1 Variables to Quantify Smart Growth Strategies

The land use variables used in this part of the research can be grouped under the four D's previously described in Chapter two. A brief description of those variables is included below. In addition, the variables used to capture TAZ-specific characteristics are also described.

4.1.1 TAZ Size Variables

The size of a TAZ is gauged in several different dimensions. The TAZ size variables we use in this research are listed in Table 4.1. Because TAZs are generally very small in town center area, and relatively big in rural area, the AREA variable is always a proxy of many other un-quantified factors. For example, bigger TAZs are associated with sparse land use, low levels of access to opportunity sites, and longer trips. Considering the fact that farm land and forecast are not associated with travel behaviors, a DAREA variable is defined by excluding forest and agricultural land uses from AREA.

Table 4.1 TAZ Size Variables Used in This Research

Variable name	Definitions
AREA	The area of a TAZ in square miles
DAREA	The developed area of a TAZ in square miles, which is obtained by subtracting forest and agricultural land uses from the total area
POP	The number of residents in a TAZ
HH	The number of households in a TAZ
TEMP	The total number of employees in a TAZ
MEMP	The number of manufacturing employees in a TAZ
REMP	The number of retail employees in a TAZ
WSEMP	The number of whole sale employees in a TAZ
OEMP	The number of all other employees in a TAZ

Population and number of households are major determinants of a TAZ's trip production. Employment is a good measure of a TAZ's attractiveness to trips. For example, in HBW trips' destination choice model, TEMP is the most important variable. In HBShop trips' destination choice, REMP explains much of the variation. More retail employment in a TAZ means stronger attractiveness of shopping trips.

4.1.2 Density

In this research, we generated two density variables: population density and employment density.

Table 4.2 Density Variables Used in This Research

Variable name	Definitions
POP_DE	POP/AREA
EMP_DE	TEMP/AREA

4.1.3 Diversity

The Diversity variables used in this research are listed in Table 4.3. Detailed descriptions of these variables were provided in Chapter 2, specifically under section 2.2.2.

Table 4.3 Diversity Variables Used in This Research

Variable name	**Definitions**
DISSM	Dissimilarity index
LD_EPY	Land use entropy
EMP_EPY	Employment entropy
NWK_LD_EPY	Non-work land use entropy
BAL	Jobs housing balance
EMPOPOP	Normalized Employment to Population Ratio

In this research, three entropy variables are defined: EMP_EPY, LD_EPY, and NWK_LD_EPY (Table 4.3). EMP_EPY uses four distinct employment types: REMP, WSEMP, MEMP, and OEMP. LD_EPY uses six land use categories as follows: residential, commercial, public, community, industrial, and park & recreation. NWK_LD_EPY uses four land use categories: residential, commercial, public, and park & recreation. These land uses are considered not to be attractors of work-related trips. Basically NWK_LD_EPY excludes two land uses from the definition of LD_EPY, which are community and industrial, because they attract a lot of work-related trips.

4.1.4 Design

The Design variables used in this research are listed in Table 4.4. Due to the limitation of data source, we generate four design variables. The first one is a binary variable: RAMP or not. It measures whether there is access to freeway in a TAZ. People who like driving would be more likely to live near freeway, so this variable could represent residents' characteristics in a TAZ. The second variable is freeway coverage rate, which is defined to be the percentage of area within 1 mile from freeway. It is another measure of freeway availability.

Road density is defined based on total area and developed area. Higher road density could improve both auto usage and non-motorized, as walk and bike paths are usually along roads. In the future study, a sidewalk variable would be more helpful in the mode selection models.

Transit coverage is defined as the proportion of area within 0.25 miles from transit stops, including both bus stops and rail stops. 0.25 miles is treated as an up-limit of walking distance for most people in Buffalo area.

Table 4.4 Design Variables Used in This Research

Variable name	**Definitions**
Ramp or not	Whether there is ramp access to freeway in a TAZ
Freeway coverage rate	Proportion of a TAZ within 1 miles from freeway
Transit coverage rate	Proportion of a TAZ within 0.25 miles from transit stops
Road density	Length of road/AREA

4.2 Trip Generation

In this step, the multiple linear regression technique is applied to relate the trip production of a TAZ to a set of zonal attributes of that TAZ. The zonal attributes considered include size variables and 4Ds variables. The 4Ds variables are explicitly considered in order to measure smart growth strategies' impacts, if any, on trip generations. Since the trip generation rates are heavily dependent on trip purposes, the trip production models are purpose specific, as is described in more detail later in this section. Finally, while the trip generation step in traditional four-step method includes both trip production and trip attraction model, the enhanced travel demand forecasting approach developed in this study does not include trip attraction models. This is because only the doubly constraint gravity model needs trip attraction, and that model is not used in the proposed framework.

4.2.1 Existing Approaches to Trip Generation Modeling

There are three major techniques developed for trip generations, including the growth factor method, multiple linear regression and cross-classification analysis (Ortuzar and Willumsen 2001). The growth factor method is heavily dependent on historical data. But smart growth strategies could change the travel behavior, making a simple extension of historical data not a good method to evaluate smart growth.

Cross-classification aims to build a multi-dimensional matrix with each dimension representing an independent variable, stratified into a number of discrete classes and categories. The contribution of each variable to trip production is explicit in this method, but the method lacks a statistical measure to assess reliability of the results. More importantly, only a limited number of variables could be included to avoid complex and degenerate cross-classification tables. That's why the cross-classification method is not selected to evaluate smart growth strategies in this study.

Multiple linear regression models address the shortcomings of both the two method mentioned above. It could incorporate many variables, and the flexibility of selecting variables makes it capable of capturing some unknown characteristics. Multiple linear regression method is used at the end.

4.2.2 Data Description

Definition of Trip Purposes

The definition of trip purpose is based on the recorded activities at the two trip ends. Table 4.5 lists the activity types available in the 2002 Buffalo-Niagara Regional Household Travel Survey and their corresponding code. Table 4.6 shows the purpose of a trip is defined according to the activities at the two trip ends. It needs to mention that Home Based School (HBSch) trip production is not modeled in this research, due to the little influence of school bus on the traffic network.

Table 4.5 Trip End Activities

Code	Description of activity
1	At home activities (eating, TV, sleeping, homemaker, etc)
2	Working at home (job related)
3	Work (include regularly scheduled volunteer work)
4	Work-related (errands, meeting, etc)
5	Attending school
6	School-related
7	Childcare
8	Quick stop (gas, ATM, coffee, etc)
9	Shopping
10	Visit friends/relatives
11	Personal business (medical/dental, errands, etc.)
12	Eat meal outside of home (restaurant, take out, etc.)
13	Entertainment, recreation, fitness
14	Civic/Religious
15	Pick up/Drop off passenger
16	Change mode of transportation
17	Other, specify

Note: This table comes from the support files of the 2002 Buffalo-Niagara Regional Household Travel Survey

Table 4.6 Definition of Trip Purposes

Destination / Origin	Home (1[1] 2)	Work (3 4)	School (5 6)	Shopping (9)	Social Recreation (10 13 14)	Other (7 8 11 12 15 16 17)
Home (1 2)		HBW	HBSch[2]	HBShop	HBSR	HBO
Work (3 4)	HBW	NHBW	NHBW	NHBW	NHBW	NHBW
School (5 6)	HBSch	NHBW	NHBO	NHBO	NHBO	NHBO
Shopping (9)	HBShop	NHBW	NHBO	NHBO	NHBO	NHBO
Social Recreation (10 13 14)	HBSR	NHBW	NHBO	NHBO	NHBO	NHBO
Other (7 8 11 12 15 16 17)	HBO	NHBW	NHBO	NHBO	NHBO	NHBO

Note: [1]: All the numbers in the parenthesis mean activity code. See Table 4.5.

[2]: Home based school

Data Weighting

The 2002 Buffalo-Niagara Regional Household Travel Survey includes a single weight variable that was developed to account for the over sampling or under sampling of particular population segments. The Census Transportation Planning Package (CTPP) data from the 2000 Census were used to calculate the household size by vehicle ownership factor. The 2000 U.S. Bureau of the Census data were used to calculate all other weight factors.

To compensate for over sampling or under sampling, the sample was balanced relative to household size and vehicle ownership by developing a weight factor (wgt1). An income weight (wgt2) was then developed to compensate for over or under representation of some income categories. Following this, a county weight (wgt3) was developed to compensate for under representation in Erie county and over representation in Niagara county. Finally, the county weight was multiplied by the product of the household size and vehicle ownership weight and income weight (wgt1*wgt2*wgt3) to produce the final weight variable (finwgt). The expansion factor (expfct) was calculated by dividing the total households based on Census 2000 data (468,719) by the number of households surveyed (2,779). When using the sample data to run population estimates, the final expansion factor was applied. This final expanded weight was the product of “finwgt” and “expfct”.

Preliminary Analysis

A preliminary data analysis is conducted to see how the trips are distributed among the different trip purposes. The trip production and weighted trip production for each trip purpose are shown in Table 4.7. While HBW trips do not have the highest market share, they are nevertheless among the most important trip purpose because they are the major reason behind congestion in morning and evening peak hours.

Table 4.7 Number of Trips Generated In Each Purpose

Trip Purpose	Total number of trips	Percentage	Total number of trips (Weighted)	Percentage
HBW	3,078	0.18	470,039	0.17
HBShop	1,358	0.08	222,304	0.08
HBSR	1,830	0.11	300,846	0.11
HBO	5,012	0.30	847,824	0.31
NHBW	933	0.06	145,787	0.05
NHBO	4,540	0.27	745,283	0.27
Sum	16,751	1	2,732,083	1

What needs to be mentioned is that the Greater Buffalo-Niagara area, where the 2002 Buffalo-Niagara Regional Household Travel Survey was conducted, includes 554 TAZs. But the parcel-level land use map we use to generate the land use variables only covers Erie county, meaning we do not have land use variables in Niagara County. As a result, all the modeling work is based on the 402 TAZs in Erie County.

4.2.3 Model Development

The data of 360 TAZs, randomly sampled from the 402 TAZs, are used as the calibration data, and the remaining 42 zones are used to validate the trip production model. Validation performance is assessed by comparing the actual production and estimated production.

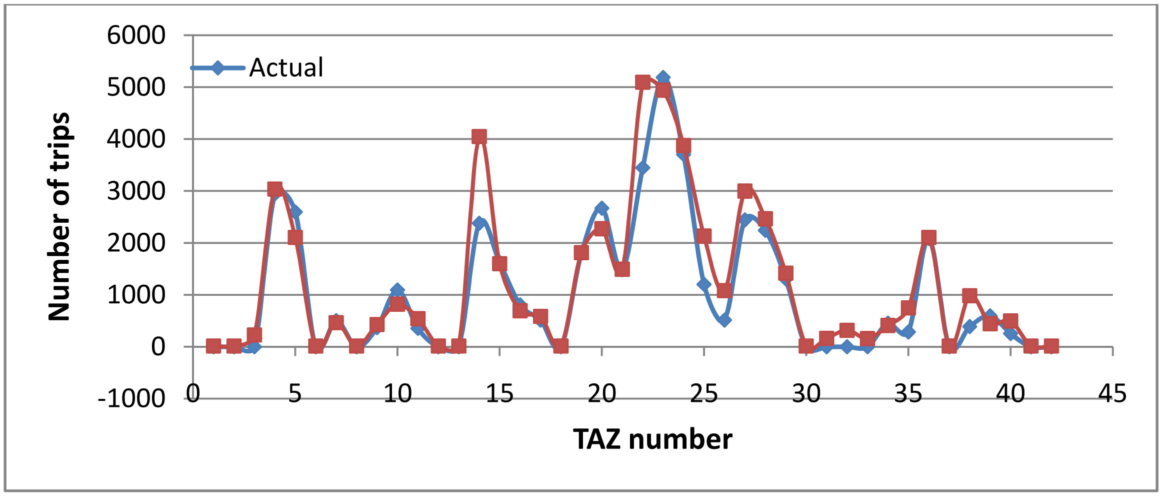
HBW Trips' Production Model

The HBW trip production model (Table 4.8) has only one independent variable: total number of workers in a TAZ, and no land use variables are found to be statistically significant in the model. Model validation results are shown in Figure 4.1, from which it can be seen that the model gives very accurate predictions.

Table 4.8 HBW Trip Production Model

Variables	**Estimated Coefficient**
Constant	16
Total number of workers	1.114***
Number of observations=360 Adjusted R squared=0.868	

Note: ***: coefficient is significant at 0.01 confidence level.
**: coefficient is significant at 0.02 confidence level
*: coefficient is significant at 0.05 confidence level



Note: The 42 TAZs are sorted by their ID and assigned values from 1 to 42 as TAZ number.

Figure 4.1 Validation Performance of HBW Trips' Production Model

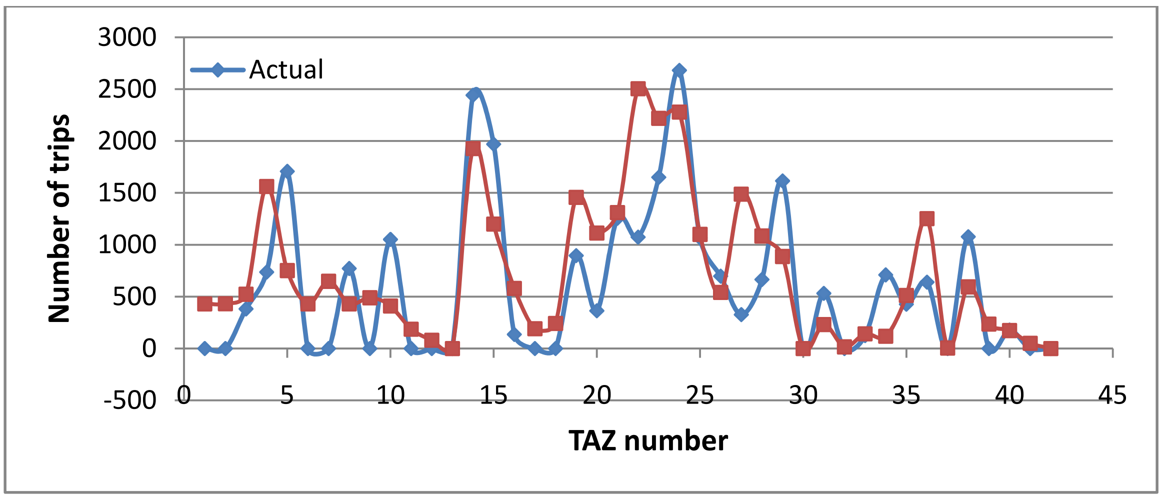
HBShop Trips' Production Model

HBShop trips' production model is shown in Table 4.9. The model has two significant independent variables: the total number of vehicles and transit coverage rate. Total number of vehicles is highly correlated with some other variables, such as population and total income of a TAZ, but it gives the highest adjusted R squared among them and is selected accordingly. The model also indicates that higher transit coverage rate promotes HBShop trips, maybe because housewives with no cars can take transit to go shopping when their husbands are not at home. Compared with HBW trips' production model, this model has lower adjusted R squared, and as a result, the validation (Figure 4.2) shows higher inconsistency.

Table 4.9 HBShop Trip Production Model

Variables	Estimated Coefficient
Constant	-115*
Total number of vehicles	0.482***
Transit coverage rate	543.6***
Number of observations=360 Adjusted R squared=0.567	

Note: ***: coefficient is significant at 0.01 confidence level.
**: coefficient is significant at 0.02 confidence level
*: coefficient is significant at 0.05 confidence level



Note: The 42 TAZs are sorted by their ID and assigned values from 1 to 42 as TAZ number.

Figure 4.2 Validation Performance of HBShop Trips' Production Model

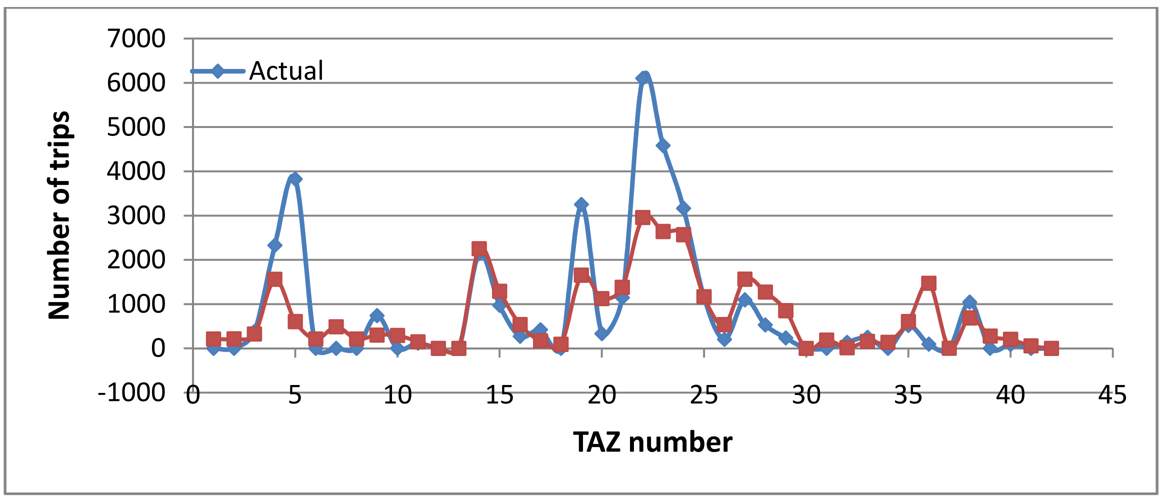
HBSR Trips' Production Model

HBSR trips' production model (Table 4.10) has exactly the same set of independent variables as HBShop model. Same with HBShop model, total number of vehicles gives the highest adjusted R squared among some correlated variables. The coefficient of transit coverage rate in this model is less than that in the HBShop model. It probably means shoppers are more likely to take transit than people who do social recreations. The model is validated (Figure 4.3), with productions of about 5 TAZs underestimated and 3 or 4 TAZs slightly overestimated.

Table 4.10 HBSR Trips' Production Model

Variables	Estimated Coefficient
Constant	-140.4**
Total number of vehicles	0.574***
Transit coverage rate	352.3***
Number of observations=360 Adjusted R squared=0.642	

Note: ***: coefficient is significant at 0.01 confidence level.
**: coefficient is significant at 0.02 confidence level
*: coefficient is significant at 0.05 confidence level



Note: The 42 TAZs are sorted by their ID and assigned values from 1 to 42 as TAZ number.

Figure 4.3 Validation Performance of HBSR Trips' Production Model

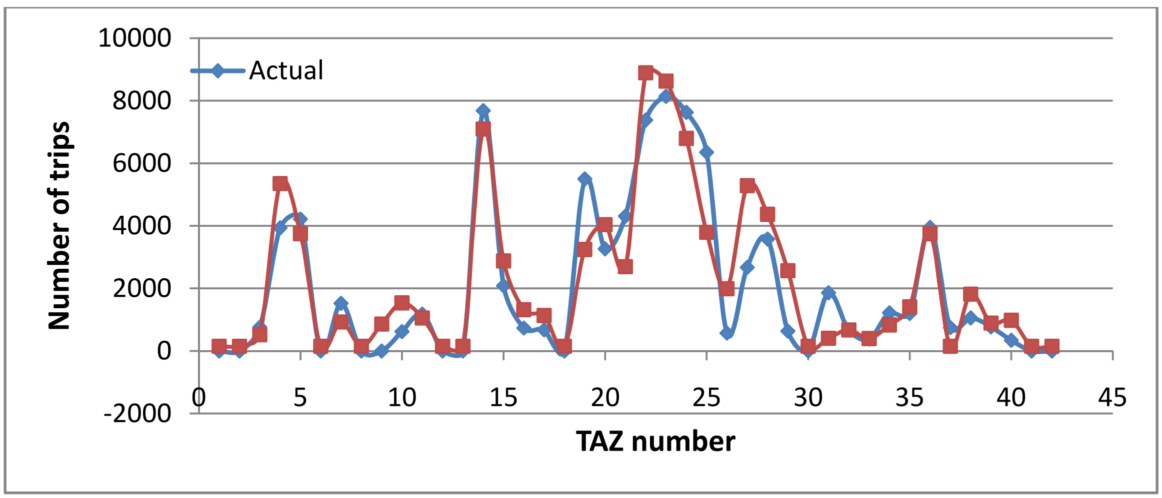
HBO Trips' Production Model

HBO trips production model (Table 4.11) has total number of workers as the only independent variable. This makes sense, because most of the "other" activities could be and more likely be done by workers, like childcare, quick stops (gas, ATM, coffee, etc), personal business, eat meal outside of home, pick up/drop off passenger and change mode of transportation. Other than childcare, all the activities are more likely to be affiliated with workers than non-workers. That's why HBO trips production is so closely related to the number of workers in a TAZ. The validation (Figure 4.4) shows excellent predictions, proving the applicability of the model.

Table 4.11 HBO Trip Production Model

Variables	Estimated Coefficient
Constant	-153.8
Total number of workers	1.9***
Number of observations=360 Adjusted R squared=0.744	

Note: ***: coefficient is significant at 0.01 confidence level.
**: coefficient is significant at 0.02 confidence level
*: coefficient is significant at 0.05 confidence level



Note: The 42 TAZs are sorted by their ID and assigned values from 1 to 42 as TAZ number.

Figure 4.4 Validation Performance of HBO Trips' Production Model

NHBW Trips' Production Model

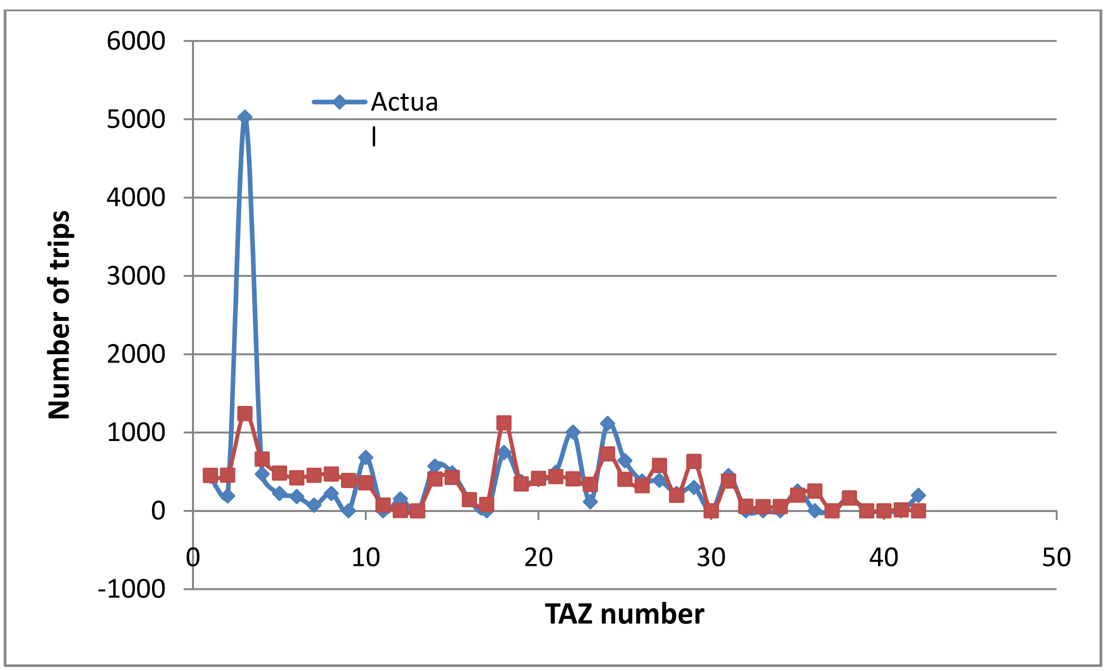
The NHBW trips' production model (Table 4.12) shows that retail employment and other employment are the two most significant independent variables. NHBW trips usually start from service facilities, like shops, restaurants, and service stores, thus the production of NHBW trips is closely related to a TAZ's retail employment. Transit service also helps a TAZ generate more

NHBW trips. The chart of validation performance (Figure 4.5) shows good predictions for all TAZs but one, which has extremely many NHBW trips. This might be because this TAZ is located in Buffalo CBD, and has too many unpredicted trips from restaurants back to work after lunch.

Table 4.12 NHBW Trips' Production Model

Variables	Estimated Coefficient
Constant	30.4
Retail employment	0.464***
Other employment	0.123***
Transit coverage rate	271***
Number of observations=360 Adjusted R squared=0.541	

Note: ***: coefficient is significant at 0.01 confidence level.
**: coefficient is significant at 0.02 confidence level
*: coefficient is significant at 0.05 confidence level



Note: The 42 TAZs are sorted by their ID and assigned values from 1 to 42 as TAZ number.

Figure 4.5 Validation Performance of NHBW Trips' Production Model

NHBO Trips' Production Model

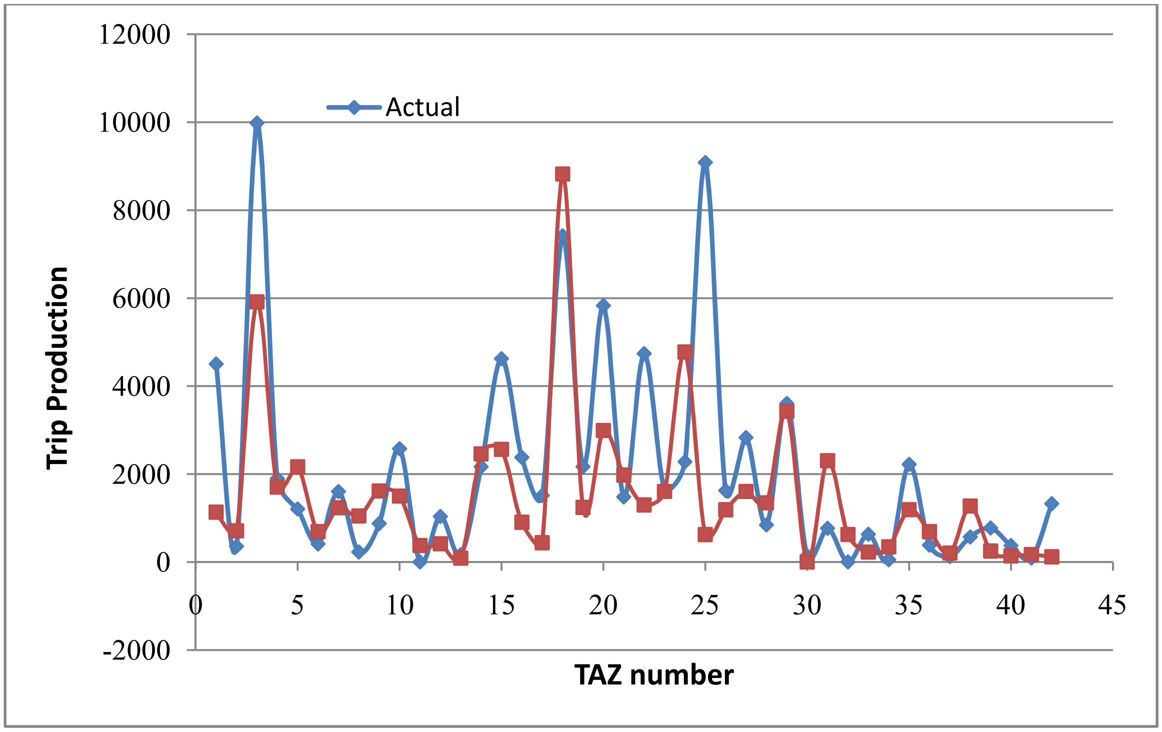
NHBO trips include all kinds of trips not classified into other trip purposes, thus have the biggest randomness, making it more difficult to be predicted by a linear regression model. The model has three significant variables (Table 4.13). It is easy to understand why retail and other employment are significant in the model, as NHBO trips are non-home-based. The total number of households in the model might be an indicator of the overall attractiveness of a TAZ. The validation performance chart (Figure 4.6) shows higher prediction errors than all other models, which can be explained by NHBO trip production's inherent randomness.

The NHB models give relatively lower R squared than the HB models. This is because the survey targeted at residents, not travelers leaving or entering a TAZ. So the actual estimation of trip production given by the survey is more accurate for HB trips than for NHB trips. That explains the lower adjusted R squared of NHB trips' production models.

Table 4.13 NHBO Trip Production Model

Variables	**Estimated Coefficient**
Constant	-20.8
Total number of HHs	0.69***
Retail employment	2.857***
Other employment	0.644***
Number of observations=360 Adjusted R squared=0.675	

Note: ***: coefficient is significant at 0.01 confidence level.
**: coefficient is significant at 0.02 confidence level
*: coefficient is significant at 0.05 confidence level



Note: The 42 TAZs are sorted by their ID and assigned values from 1 to 42 as TAZ number.

Figure 4.6 Validation Performance of NHBO Trips' Production Model

4.3 Destination Choice Models

With the trip generation or production models developed, the next step is that of trip distribution. In the enhanced four-step travel demand forecasting method developed in this study, the trip distribution step is accomplished using destination choice models which estimate Origin-Destination (O-D) demand for different trip purposes. Various discrete choice modeling techniques were applied to quantify the relationship between individual trips' destination choice decisions and the potential affecting attributes using the 2002 Buffalo-Niagara Regional Household Travel Survey data. The attributes that were taken into consideration included: (1) travelers' socio-economic characteristics such as age, gender, and household structure; (2) trip related attributes such as travel distance and travel time; (3) transportation network related attributes such as the availability of highways to the origin zone or an alternative destination zone of a trip; and (4) *land use variables such the 4D indices of a destination zone*. The inclusion of the land use variables, particularly the ones related to the destination alternatives, allows the enhanced forecasting method to quantify how land use patterns affect travelers' destination choice behavior.

In addition, since the destination choice decision heavily rely on what a trip is made for, destination choice models are developed for five trip purposes respectively, which are home based work (HBW) trips, home based shopping (HBShop) trips, home based social recreational and other (HBSRO), non-home based work (NHBW) trips, and non-home based other (NHBO) trips. The major outcomes of this section are the best destination choice models that can be used to predict the probability for an individual trip with certain purpose to choose a zone as the travel destination. These destination choice probabilities are then aggregated to estimate the OD trip tables of different trips purposes.

4.3.1 Existing Approaches to Trip Distribution Modeling

There are three major types of trip distribution models that are used to predict OD demand, including: (1) growth-factor methods; (2) gravity models; and (3) destination choice models. Among them, the first two are aggregate models for estimating the number of trips between each OD pair while the last one is disaggregate models that focus on destination choice decisions of individual trips (Ortuzar and Willumsen 2001) The concepts, advantages and disadvantages of them are discussed below.

Growth factor methods estimate the future-year OD demand by inflating the base-year OD demand by certain ratios (called growth factors). There are different variants such as the uniform growth-factor methods, the singly constrained growth-factor methods, and the doubly constrained growth-factor methods, depending on if a uniform growth factor or more sophisticated growth factors are used. These methods do not consider the impact of transportation costs on future travel demand, and thus are little sensitive to policies (such as new land use patterns and new transportation modes) that may significantly improve the system performance, In this context, they are not appropriate to evaluate smart growth strategies.

Gravity models are the most commonly used trip distribution models. They usually share the following basic functional form as follows (Ortuzar and Willumsen 2001):

$$T_{ij} = A_i O_i B_j D_j f(c_{ij})$$

Where

T_i=Total number of trips produced from TAZ *i*;

O_j=Trip production of TAZ *i*;

D_j=Trip attraction of TAZ *j*;

$f(c_{ij})$=Friction factor which is usually a decreasing function of the travel cost between TAZ *i* and TAZ *j*;

$A_i = {}^{1}\!\big/_{\sum_j B_j D_j f(c_{ij})}$

$B_i = {}^{1}\!\big/_{\sum_j A_i O_i f(c_{ij})}$

Several disadvantages of gravity models prevent them from being used to evaluate smart growth strategies. For example, these models only use automobile travel times as OD travel cost, ignoring travel times of other modes such as walking, biking and transit. Moreover, gravity models do a bad job modeling intra-zonal trips or local travel within a given TAZ. For instance, smart growth strategies place an emphasis on mixed land uses, and walkable neighborhoods. Such higher non-motorized connectivity, higher development density and higher land use mixtures are not well represented by relatively large TAZs and zone centroids connectors. (DKS_Associates 2007). As a consequence, gravity models are indifferent to the smart-growth neighborhood designs and the designs of conventional suburban neighborhoods with disconnected local networks and homogenous land-uses. What's more, since gravity models focus on the distribution trends of trips but not individual trip makers, they cannot capture the impact of new policies such as smart growth on travelers' travel decisions.

In contrast to the aforementioned approaches, destination choice models are disaggregate and individual-trip oriented. These models focus on the choice situation in which a traveler faces all the TAZ in an area as potential destination alternatives when making a trip from an origin TAZ. They estimate the probability of a traveler choosing a destination alternative, assuming that this probability is mainly affected by the corresponding traveler's socio-demographic attributes, travel cost, land use characteristics of the destination zone, and so on. Among many potential affecting factors, land use characteristics of a destination TAZ can be measured by the quantitative 4D variables (Density, Diversity, Destination and Design). These land use variables, when combined, serve as a generalized attractiveness indicator for a destination TAZ. Different from the double-constrained gravity models, which consider both the trip production and the trip attraction of a zone as modeling constraints, the constraint of trip attractions are relaxed in destination choice models. It means travelers have more freedom to choose their favorite destinations, thus giving the model more flexibility to reflect potential travel pattern changes resulted from smart growth.

As there are always hundreds or thousands of TAZs in an area, and it is not practical to estimate a model with so many alternatives, one major challenge of destination choice modeling is to reduce the size of the alternative set. To do so, many researchers used random sampling method to get fewer representatives of destination alternatives. For instance, Bhat et al. randomly selected 9 zones plus the actual chosen zone from 986 TAZs in the Boston Metropolitan Area as an alternative set in a study of home based social recreation trips' destination choice (Bhat, Govindarajan et al. 1998). The San Francisco's travel demand model has 40 stratified sampled alternatives out from 1728 TAZs (Jonnalagadda, Freedman et al. 2001). The Boise, Idaho model's destination choice part was also estimated by using a stratified importance sampling. The sample includes origin zone, the destination zone, a sample of 20 out of 40 zones nearest to, but not including, the origin, and 20 zones of the remaining zones. Thus it sampled 42 alternatives from several hundred TAZs in total (Shiftan 1998). New Hampshire model uses very similar number of alternative (Cambridge_Systematics_Inc 1998). What these researchers typically do is to firstly group the original destination alternatives into a number of strata, and then sample certain number of destination alternatives from each stratum based on the importance level of this stratum. These sampled destination alternatives, when grouped, formed the new destination choice set (Ben-Akiva and Lerman 1985). In Tel Aviv, the full set of 1244 zones was used for the destination-choice models (Shiftan and Ben-Akiva 2010). While most models predict location choices at the TAZ level, the Sacramento model predicts location at the parcel level. A sample of 100 parcels from more than 700,000 was used for estimation (Shiftan and Ben-Akiva 2010).

There are many applications of destination choice models, particularly in the activity-based travel demand forecasting paradigm (Cambridge_Systematics_Inc 1998; Shiftan 1998; Jonnalagadda, Freedman et al. 2001). Destination choice models may also be combined with other travel choice models. For example, in the San Francisco model, destination choices and mode choices are modeled simultaneously, by using a nested choice tree structure (Jonnalagadda, Freedman et al. 2001). Greenwald built several binary logistic choice models to estimate the decisions to travel intrazonally or interzonally, and reached the conclusion that intrazonal trips might be influenced by urban form (Greenwald 2006). Bhat built a Multinomial Logit model in a research to capture destination choices of home-based social recreation trips. The modeling results showed that land use attributes, such as a zone's retail space and non-retail space, and percentage of water area, could influence the attractiveness of a zone for recreational activities (Bhat, Govindarajan et al. 1998). In the San Francisco destination choice model, a destination zone's characteristics are measured by including some dummy variables indicating whether a zone is in CBD, UBD, Silicon Valley, etc (Jonnalagadda, Freedman et al. 2001). These applications provide rich information about the types of affecting factors that should be considered in the current study.

4.3.2 Model Formulation

In the specific destination choice situation, each trip that originates in a TAZ can choose any TAZ (including the origin TAZ) as the destination. Thus, all the zones in the Buffalo region (402 TAZs in total) can theoretically be considered as the destination alternatives for a trip. In order to reduce the size of the alternative set, we applied a stratified sampling technique (Ben-Akiva and Lerman 1985). Following the sampling process, we firstly stratified all the 402 TAZs to four strata, which are: (1)stratum 1 that only includes origin TAZ; (2) stratum 2 that includes all the other TAZs that are less than d_1 miles away from the origin TAZ; (3) stratum 3 that includes the TAZs in the distance between d_1 and d_2 miles away from the origin TAZ; and (4) stratum 4 that includes all the remaining TAZs (see Figure 4.7).

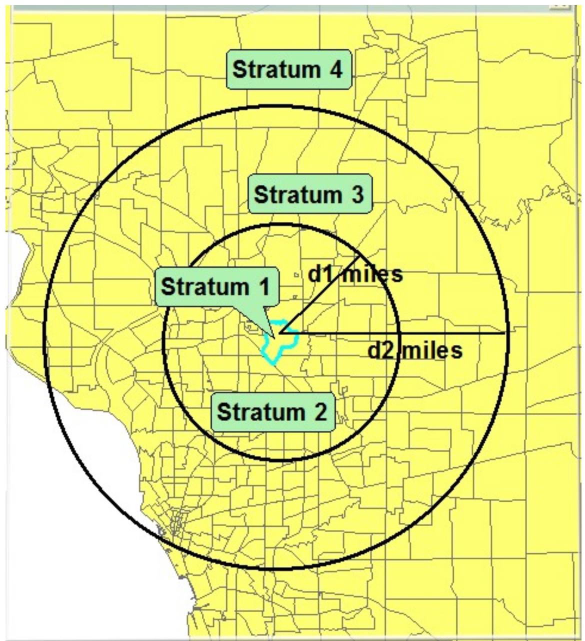


Figure 4.7 Illustration of Destination Choice Sampling Strategy

After the stratification, we sampled one TAZ from each stratum as the new destination alternatives. Therefore, the new destination choice set includes four TAZs. It needs to mention that the chosen destination TAZ is always included in the new choice set, and thus no TAZ was sampled from the stratum in which the chosen TAZ is included. The boundaries of the strata are chosen using the 33% percentile and 66% percentile, so that strata 2, strata 3 and strata 4 have the same number of chosen alternatives. The values of d_1 and d_2 are given together with the modeling results later in this section.

Given such a stratified sampling process, the decision to choose alternative i among the sampled choice set as the destination can be represented by a enhanced Multinomial Logit model as shown below:

$$\boldsymbol{P_n(i|D)} = \frac{\frac{1}{q_{in}} e^{\mu^* \overline{V_{in}} + \mu^* \ln(M_i) + \mu' \ln(B_{in})}}{\sum_{j \epsilon D} \frac{1}{q_{jn}} e^{\mu^* \overline{V_{jn}} + \mu^* \ln(M_j) + \mu' \ln(B_j)}}$$

Where:

$P_n(i|D)$=The conditional probability of alternative i being chosen given the choice set D for trip n;

q_{in}=The selection probability of sampling the desired number of alternatives from the stratum where alternative i belongs to, for trip n. It can be used as the expansion factor for alternative i;

$\overline{V_{in}}$=The expected utility of choosing alternative i for n;

M_i=A size measure of the stratum where alternative i belongs to;

B_{in}=The measure of the variability of the utilities of alternatives in the stratum where alternative i is sampled from;

$\mu' = {}^{\mu^*}/_{\mu}$- The nonnegative ratio of the corresponding scale parameters;

For alternatives within the same stratum, the correlation coefficient between them is equal to $(1 - \mu'^2)$ Therefore, μ' needs to satisfy: $0 \leq \mu' < 1$. If μ' equals to 1, the correlation coefficient between alternative utilities is zero. In this case, the parameters of the specified choice model are not dependent on the definition of strata. Hence, μ' can be seen as a measure of the stratification strategy's efficiency.

B_{in}, as a measure of the variability of the choice utilities in a stratum, is usually difficult to be observed or modeled. Considering this, it was assumed that B_{in} is approximately equal among strata. Therefore, the equation is simplified as follows.

$$\boldsymbol{P_n(i|D)} = \frac{\frac{1}{q_{in}} e^{\mu^* \overline{V_{in}} + \mu^* \ln(M_i)}}{\sum_{j \epsilon D} \frac{1}{q_{jn}} e^{\mu^* \overline{V_{jn}} + \mu^* \ln(M_j)}}$$

The size measurement chosen for this study is total employment, retail employment, population, number of households, number of TAZs, and area. Each size variable was tested when we built the model and the best was selected.

4.3.3 Data Sources and Data Assembly

The data source for this study is the 2002 Buffalo-Niagara Regional Household Travel Survey. The survey collected data on socio-demographic attributes of the household. The survey also included a one-day activity diary to be filled by all members of the household more than 5 years old. The survey recorded more than 20,000 trips.

There are 551 TAZs in the Greater Buffalo-Niagara area. Unfortunately, GBNRTC only provided a more detailed parcel map for Erie County, which includes 402 TAZs, as mentioned before. Since the generation of more detailed land use attributes, such as diversity indices, relies on the parcel map information, all the modeling work is based on the 402 TAZs of Erie County but not Niagara County. GBNRTC also provided a TransCAD TAZ layer and transportation network layer of the whole region, from which we got basic land use attributes of each TAZ, such as area, population, number of households, total employment, retail employment, average household income, average household size, etc. We also generated the travel distances and travel times of different modes for each OD pair based on the TransCAD network distances, reported travel times in the survey, and the average speeds of different modes. These attributes, when combined, form the pool of the potential affecting factors for the discrete choice modeling.

1. The steps to generate the modeling data are as follows:
2. The TAZ-based land use attributes were created from the information available in the parcel land use map;

3. The trips recorded in the travel survey were categorized into five groups by trip purposes, including HBW trips, HBShop trips, HBSRO trips, NHBW trips, and NHBO trips;
4. For the data set of each trip purpose, the origin and destination of a trip were projected to the TransCAD TAZ layer and traffic network map. The shortest network distance for each trip was generated using a GIS tool. All the trips of each purpose, excluding intrazonal trips, were sorted by trip distances. Then the two distance cutpoints d_1 and d_2 that are used to define the destination alternative sampling strata were determined so that the trips were divided into three equal-size groups. We have five trip purposes, thus different sets of distance cutpoints were generated for different trip purposes;
5. For each trip of a purpose, four candidate destination TAZs were selected based on the distance cutpoints. To be more specific, origin TAZ itself was alternative 1. Alternative 2 was randomly sampled from the TAZs which are less than d_1 miles from the origin TAZ. Alternative 3 was randomly sampled from the TAZs that are within d_1 miles and d_2 miles from the origin TAZ. At last, alternative 4 was randomly sampled from the TAZs that are further than d_2 miles away from the origin TAZ. The chosen TAZ was selected automatically, and no TAZ was sampled from the stratum where the chosen TAZ is; and
6. The travel survey trip files were combined with the corresponding person/household socio-demographic files and TAZ level land use files, using SQL application in ACCESS, to append the socio-demographic characteristics of the individual to each of the trip production, and land use attributes to each of the trip ends.

The final data included 1334 trip observations for the HBW purpose, 679 trip observations for the HBShop purpose, 3035 trip observations for the HBSRO purpose, 933 trip observations for the NHBW purpose, and 4113 observations for the NHBO purpose. 80% of the observations in each purpose were used to calibrate the model, and the rest 20% were used for model validation.

4.3.4 Modeling Results

Destination Choice Model of Home Based Work Trips

There are 1,002 HBW trip observations used for model calibration, and 332 trip observations for model validation. The distance cut-points d_1 and d_2 are 4.65 miles and 9 miles, which were the 33 percentile and 66 percentile distances of the trip distribution respectively (see Figure 4.8). A number of models were specified, and the best model is shown in Table 4.14. This best model starts with a base model that only includes socio-demographic variables and trip distances. After the inclusion of land use variables to the base model, the MacFadden Pseudo R squared goes up by 50% from 0.22 to 0.33. This indicates that land use variables play significant roles in explaining travelers' destination choice behavior.

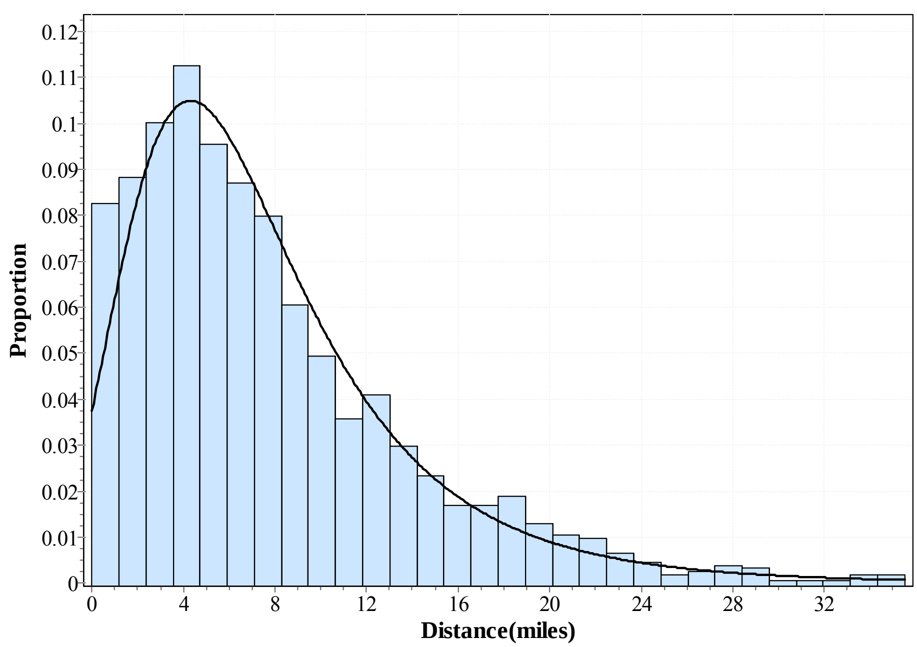


Figure 4.8 Trip Length Distribution of HBW Trips

Table 4.14 Destination Choice Model of HBW Trips

Variable	Alt. 1	Alt. 2	Alt. 3	Alt. 4
Alternative specific constant		-0.8657 (0.000)[1]	-1.1022 (0.000)	-1.6884 (0.000)
Piecewise distance between 0~1.5 miles	-0.02876 (0.000)	-0.02876 (0.000)	-0.02876 (0.000)	-0.02876 (0.000)
Piecewise distance between 1.5~3 miles	-0.5385 (0.000)	-0.5385 (0.000)	-0.5385 (0.000)	-0.5385 (0.000)
Piecewise distance between 3~5 miles	-0.3691 (0.000)	-0.3691 (0.000)	-0.3691 (0.000)	-0.3691 (0.000)
Piecewise distance more than 5 miles	-0.1426 (0.000)	-0.1426 (0.000)	-0.1426 (0.000)	-0.1426 (0.000)
Household Income of traveler (1000 dollars)	-0.01245 (0.000)			
Female or not: 1 if yes, 0 if not.		-0.1191 (0.000)	-0.1191 90.000)	-0.5771 (0.000)
Traveler's age	0.01048 (0.000)			
Number of jobs participated by the traveler	0.6865 (0.000)			
Activity duration		0.0005408 (0.000)	0.002075 (0.000)	0.004479 (0.000)
Total employment of destination TAZ	0.0001612 (0.000)	0.0001612 (0.000)	0.0002260 (0.000)	0.0003568 (0.000)
Dissimilarity index of destination TAZ	0.5703 (0.000)	0.4802 (0.000)		
Land use entropy of destination TAZ	0.6652 (0.000)	0.6652 (0.000)		
Normalized employment-to-population ratio	0.1629 (0.000)	0.6071 (0.000)	0.6071 (0.000)	0.6071 (0.000)

Freeway coverage rate	-0.6792 (0.000)			
Log(Total employment of all TAZs in a stratum)	0.5949 (0.000)	0.5949 (0.000)	0.5949 (0.000)	0.5949 (0.000)
Summary statistics Number of observations =1002 Base model's McFadden Pseudo R squared=0.2221 Final model's McFadden Pseudo R squared after adding land use variables=0.3307				

Note: [1]: The value in the parentheses is the p-value of the coefficient.

As shown in the best model, 16 variables, including land use attributes play significant roles in explaining the destination choice behavior of HBW trips. Travel distance enters the utility of all the four groups in a piecewise fashion to take into account a possible nonlinear marginal impact of distance on travelers' destination choices. And the results show exactly the case: within 1.5 miles, distance has relatively small marginal impact (-0.008); when the distance is greater than 1.5 miles, the marginal impact of it becomes stronger but then decreases as distance gets larger. Such a pattern is consistent with the trip frequency distribution (Figure 4.8). Higher household income reduces the possibility of selecting alternative 1 (i.e., intrazonal travel), and this might be a result of mid-class residential suburbanization. Mid-class families tend to live in suburban neighborhoods, which are further away from job locations, and a result their HBW trips are more likely to be inter-zonal. Females and older people make more intra-zonal trips than males and younger people. It is interesting to find that having more jobs encourages intrazonal travels. This could be explained by time constraints: if a person needs to pursue multiple jobs, the most efficient way is to arrange jobs near home so that s/he does not need to spend too much time in traveling. Activity duration is positively related to the utilities of alternative 2, 3 and 4 (interzonal travels). It might be because longer time spent at work makes a long travel worthwhile.

In terms of land use variables, we can see that more employment opportunities attract more HBW trips. In addition, the employment-to-population ratio exerts extra positive effects on all the four utilities. The positive values of two diversity-related land use variables (land use entropy and dissimilarity index) in utility function 1 and utility function 2 also make TAZs in these two groups more attractive. In addition, the parameter of dissimilarity index is higher for alternative 1 than that for alternative 2, indicating that intra-zonal travels are more likely to be induced by dissimilar land uses than inter-zonal travels. Moreover, the negative coefficient of the freeway coverage rate (which represents the proportion of a destination TAZ which have freeway in one mile away) in alternative 1 show that easy access to freeway actually discourages intra-zonal travels. This finding supports that idea that highway facilities induce more auto-oriented long distance travels.

Table 4.15 Validation Performance of the HBW Trips' Destination Choice Model

HBW Trips	**Alt.1**	**Alt.2**	**Alt.3**	**Alt.4**	**Overall**
Actual	24	104	100	104	332
Estimated	4	108	104	116	332
Correctly estimated	3	63	62	77	205
Percentage of correctly estimated	12.5%	60.6%	62%	74.0%	61.7%

The model was validated using 332 trip observations. The validation results in Table 4.15 shows that the overall prediction accuracy is 61.7%. In general, the model tends to overestimate the probability of TAZs further away while underestimating the possibility of choosing nearer TAZs. The reason could be that the true destination choice behavior, particularly for intrazonal trips, is not well revealed by the modeling data, due to the fewer intrazonal trips available in the modeling dataset.

Destination Choice Model of Home Based Shop Trips

There are 510 HBShop trip observations in the calibration data, and 169 observations in the validation data. The distance cut-points d_1 and d_2 are 2.28 miles and 4.5 miles, which represent the 33 percentile and 66 percentile distances of the trip distribution respectively as shown in Figure 4.9. A number of models were specified, and the best model is shown in Table 4.16. This best model starts with a base model that only includes socio-demographic variables and trip distances. After the inclusion of land use variables to the base model, the MacFadden Pseudo R squared goes up from 0.29 to 0.48 by 65.5%. That means land use variables play significant roles in explaining travelers' destination choice behavior for HBShop trips as well.

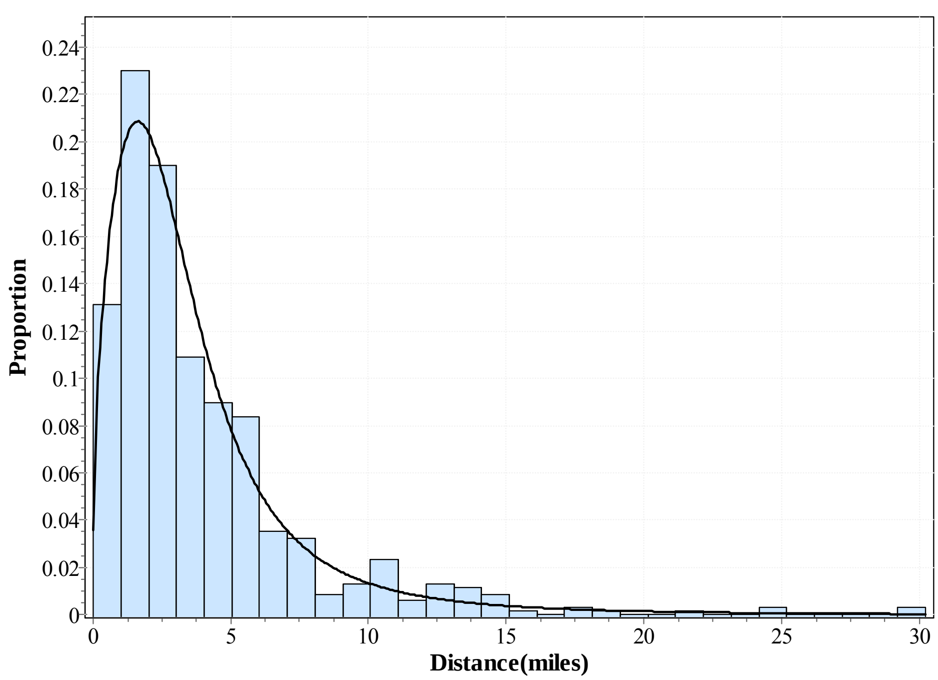


Figure 4.9 Trip Length Distribution of HBShop Trips

Table 4.16 Destination Choice Model of HBShop Trips

Variable	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**
Constant		-1.1558 (0.000)[1]	2.0191 (0.000)	-0.3754 (0.000)
Distance	-1.5869 (0.000)	-0.3169 (0.000)	-1.0301 (0.000)	-0.3092 (0.000)
Activity duration	-0.03016 (0.000)	-0.005427 (0.000)	-0.00174 (0.000)	
Female or not: 1 if yes, 0 if not	0.4869 (0.000)	0.3392 (0.000)	0.3345 (0.000)	
Population of DTAZ	0.000141 (0.000)	0.000141 (0.000)	0.000141 (0.000)	0.000141 (0.000)
Retail employment of Destination TAZ	0.9008E-4 (0.000)	0.001451 (0.000)	0.001308 (0.000)	0.001842 (0.000)
Land use entropy of Destination TAZ	1.1910 (0.000)	0.0013 (0.000)		
Transit coverage rate of Destination TAZ	2.5217 (0.000)	2.5217 (0.000)	0.4114 (0.000)	
Log(Number of retail employees of all TAZs in a group)	0.5579 (0.000)	0.5579 (0.000)	0.5579 (0.000)	0.5579 (0.000)
Summary statistics Number of observations =510 Base model's McFadden Pseudo R squared=0.2930 Final model's McFadden Pseudo R squared after adding land use variables=0.4791				

Note: [1]: The value in the parentheses is the p-value of the coefficient.

As shown in Table 4.16, nine types of variables, including land use variables influence the destination choice behavior of HBShop trips. Trip distance is the most important factor and enters the four utility functions directly. The distance parameter in alternative 1 is the largest among all the four distance parameters, meaning that a traveler is more reluctant to travel intrazonally than interzonally with the increase of travel distance. Longer activity duration encourages the selection of alternative 4, and this might be a result of the isolation of large shopping centers from residential areas. Females prefer nearer destinations more than males, as indicated by its negative parameters in alternative 2, 3 and 4.

In terms of land use variables, a HBShop trip is more likely to visit a destination with more retail employment. The positive parameters of land use entropy in utility function 1 and function 2 indicate that higher land use entropy makes a TAZ more attractive to HBShop trips, and the affects go to zero for the TAZ further away. Moreover, TAZs with higher transit coverage could attract more HBShop trips, but the attractiveness drops down to zero if the TAZ is more 4.5 miles away.

Table 4.17 Validation Performance of the HBShop Trips' Destination Choice Model

HBShop Trips	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Sum**
Actual	33	46	45	45	169
Estimated	21	68	41	39	169
Correctly Estimated	12	39	28	35	114
Prediction Accuracy	36.4%	84.8%	62.2%	77.8%	67.5%

Table 4.17 shows the validation results of the model. Again, the model tends to overestimate the probability of TAZs further away while underestimating the possibility of choosing nearer TAZs. The possible reason could be that the true destination choice behavior, particularly for intrazonal trips, is not well revealed by the modeling data, due to the fewer intrazonal trips available in the modeling dataset.

Destination Choice Model of Home Based Social Recreation and Home Based Other Trips

There are 2,421 HBSRO trip observations in the calibration data, and 614 observations in the validation data. The distance cutpoints d_1 and d_2 are 2.64 miles and 5.08 miles, which are the 33 percentile and 66 percentile distances of the trip distribution respectively as shown in Figure 4.10.

A number of models were specified, and the best model is shown in Table 4.18. This best model starts with a base model that only includes socio-demographic variables and trip distances. After the inclusion of land use variables to the base model, the MacFadden Pseudo R squared goes up from 0.17 to 0.21 by 23.5%.

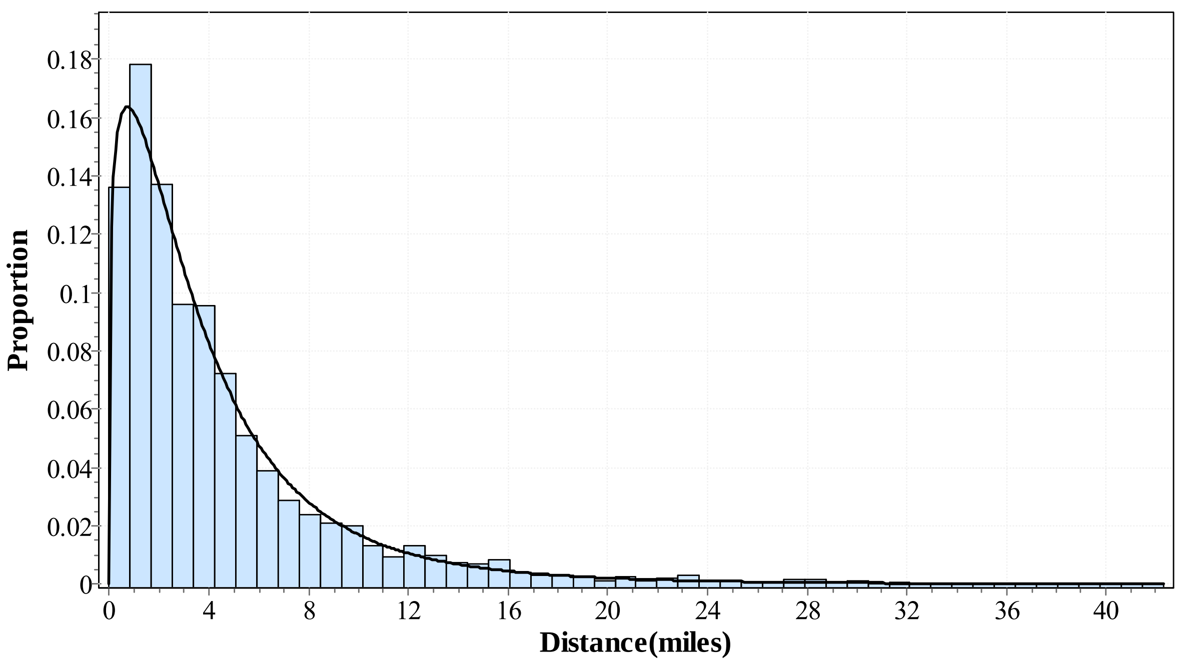


Figure 4.10 Trip Length Distribution of HBSRO Trips

Table 4.18 Destination Choice Model of HBSRO Trips

Variable	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**
Constant		0.2997 (0.000)[1]	0.1687 (0.000)	-1.2654 (0.000)
Distance	-0.9233 (0.000)	-0.8216 (0.000)	-0.5368 (0.000)	-0.2079 (0.000)
Female or not: 1 if yes, 0 if not.	0.1926 (0.000)			
License or not: 1 is yes, 0 if not	-0.3631 (0.000)			
Activity duration		0.001507 (0.000)	0.005031 (0.000)	0.007236 (0.000)
Total households of Destination TAZ	0.0003703 (0.000)	0.0003703 (0.000)	0.0001941 (0.000)	0.0001941 (0.000)
Other employment of Destination TAZ	0.2008E-4 (0.000)	0.2008E-4 (0.000)		
Freeway coverage rate		0.06183 (0.000)	0.01418 (0.000)	0.01418 (0.000)
Log(Total employment of all TAZs in a group)	0.5706 (0.000)			
Summary statistics Number of observations =2421 Base model's McFadden Pseudo R squared=0.1743 Final model's McFadden Pseudo R squared=0.2082				

Note: [1]: The value in the parentheses is the p-value of the coefficient.

As is shown in Table 4.18, nine variables, including land use variables, influence the destination choice behavior of HBSRO trips. Two size-related land use variables, the number of households and other employment, and one Design-related land use variable, Freeway coverage rate, are main factors that attract HBSRO trips. In terms of socio-demographic variables, females prefer nearer destinations, and licensed people prefer further destinations. Longer activity duration increases the likelihood of making longer trips. The R squared of HSRO model is relatively low. This could be explained by the internal chaos of HBSRO trips, which include all trips that cannot be categorized into other trip purposes.

Table 4.19 Validation Performance of the HBSRO Trips' Destination Choice Model

HBSRO Trips	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Overall**
Actual	112	205	120	177	614
Estimated	50	246	153	165	614
Correctly Estimated	26	147	70	117	360
Prediction Accuracy	23.2%	71.7%	58.3%	66.1%	58.6%

Table 4.19 shows the validation results of the model. The overall prediction accuracy is 58.6%. Again, the model tends to overestimate the probability of TAZs further away while underestimating the possibility of choosing nearer TAZs. The possible reason is the same with those of HBW model and HBShop model.

Destination Choice Model of Non Home Based Work Trips

There are 748 NHBW trip observations in the calibration data, and 185 observations in the validation data. The distance cut-points d_2 and d_2 are 2.53 miles and 5.6 miles, which represent the 33 percentile and 66 percentile distances of the trip distribution respectively as shown in Figure 4.11. The trip distribution figure shows that NHBW trips have shorter than other trip purposes. This could be explained by trip chaining behavior. A HBW trip chain could include at least one HBShop or HBSRO trip, and a NHBW trip. In this case NHBW trips are always shorter than HBW trips. NHBW could also be a trip back to work after lunch break. The lunch trip is more likely shorter than the distance from home to work.

A number of models were specified, and the best model is shown in Table 4.20. This best model starts with a base model that only includes socio-demographic variables and trip distances. After the inclusion of land use variables to the base model, the MacFadden Pseudo R squared goes up from 0.19 to 0.27 by 42.1%. That means land use variables play significant roles in explaining travelers' destination choice behavior.

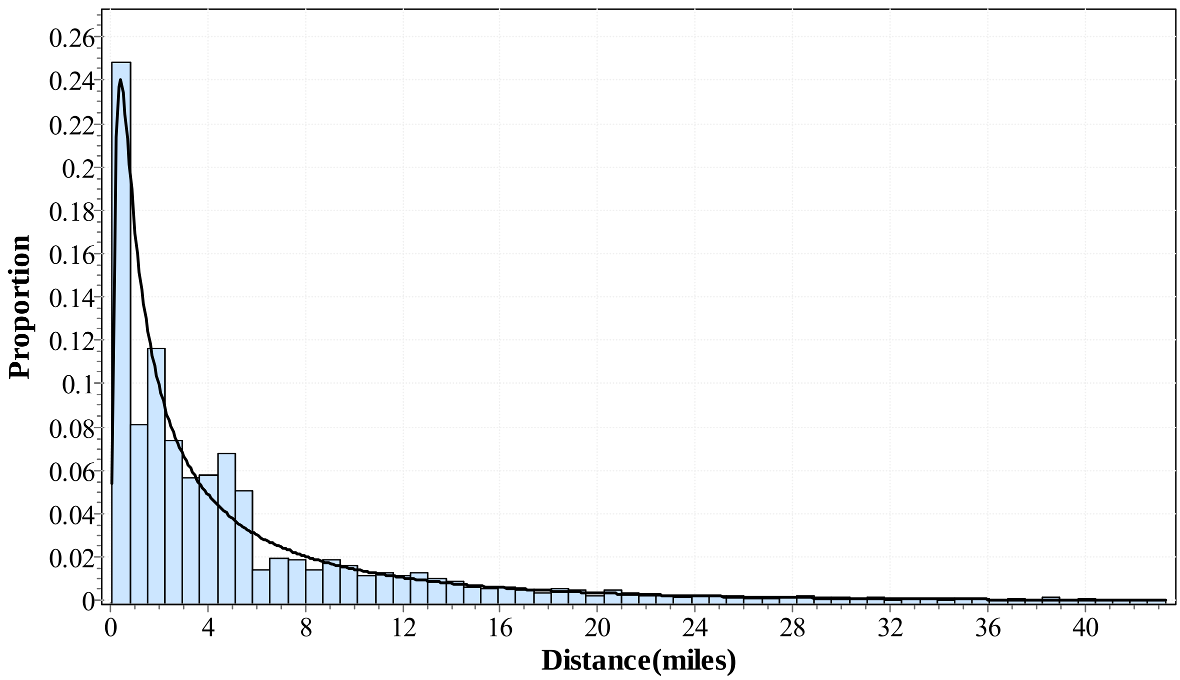


Figure 4.11 Trip Length Distribution of NHBW Trips

Table 4.20 Destination Choice Model of Non Home Based Work Trips

Variable	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**
Constant		1.1652 (0.000)[1]	1.6801 (0.000)	1.1497 (0.000)
Distance	-0.4591 (0.000)	-0.6164 (0.000)	-0.4186 (0.000)	-0.1858 (0.000)
Female or not: 1 if yes, 0 if not.		-0.5635 (0.000)	-0.5635 (0.000)	-1.1679 (0.000)
Activity duration	-0.002629 (0.000)			
Traveler's age	0.01969 (0.000)			
Total employment of Destination TAZ	0.0001144 (0.000)	0.0002614 (0.000)	0.0003175 (0.000)	0.0004489 (0.000)
Non work land use entropy of Destination TAZ	0.9465 (0.000)	0.9465 (0.000)		
Dissimilarity index of Destination TAZ	0.2202 (0.000)	0.2202 (0.000)		
Log(Total employment of all TAZs in a group)	0.5374 (0.000)	0.5374 (0.000)	0.5374 (0.000)	0.5374 (0.000)
Summary statistics Number of observations =748 Base model's McFadden Pseudo R squared=0.1910 Final model's McFadden Pseudo R squared=0.2686				

Note: [1]: The value in the parentheses is the p-value of the coefficient.

Similar with the previous models, the NHBW trips' destination choice model has variables like trip distance, gender, age and activity duration. Females and/or senior citizens prefer nearer TAZs, and longer activity duration increases the utility of further TAZs.

In terms of land use variables, a NHBW trip is more likely to visit a destination with more total employment, and the same amount of total employment seems to have stronger attraction power in a TAZ further away. Non work land use entropy and dissimilarity index are significant in the utility functions of alternative 1 and alternative 2, indicating that shorter travels are more likely to be induced by diversified land uses than longer travels.

Table 4.21 Validation Performance of the NHBW Trips' Destination Choice Model

NHBW trips	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Overall**
Actual	18	52	53	62	185
Estimated	7	47	55	76	185
Correctly Estimated	3	29	27	47	106
Prediction Accuracy	16.7%	55.8%	50.9%	75.8%	57.3%

The model validation is shown in Table 4.21. Again with the same reason with other models, the model tends to overestimate the probability of TAZs further away while underestimating the possibility of choosing nearer TAZs.

Destination Choice Model of Non Home Based Other Trips

There are 3,302 NHBO trip observations in the calibration data, and 811 observations in the validation data. The distance cutpoints d_2 and d_2 are 2.55 miles and 5.35 miles, which are the 33 percentile and 66 percentile distances of the trip distribution respectively as shown in Figure 4.12. A number of models were specified, and the best model is shown in Table 4.22. This best model starts with a base model that only includes socio-demographic variables and trip distances. After the inclusion of land use variables to the base model, the MacFadden Pseudo R squared goes up from 0.14 to 0.18, by 28.6%. That means land use variables play moderate roles in explaining travelers' destination choice behavior. NHBO trips have great internal chaos, as they include all trips that could not be categorized into other trip purposes. That's why the R squared of this model is the lowest among the five destination choice models.

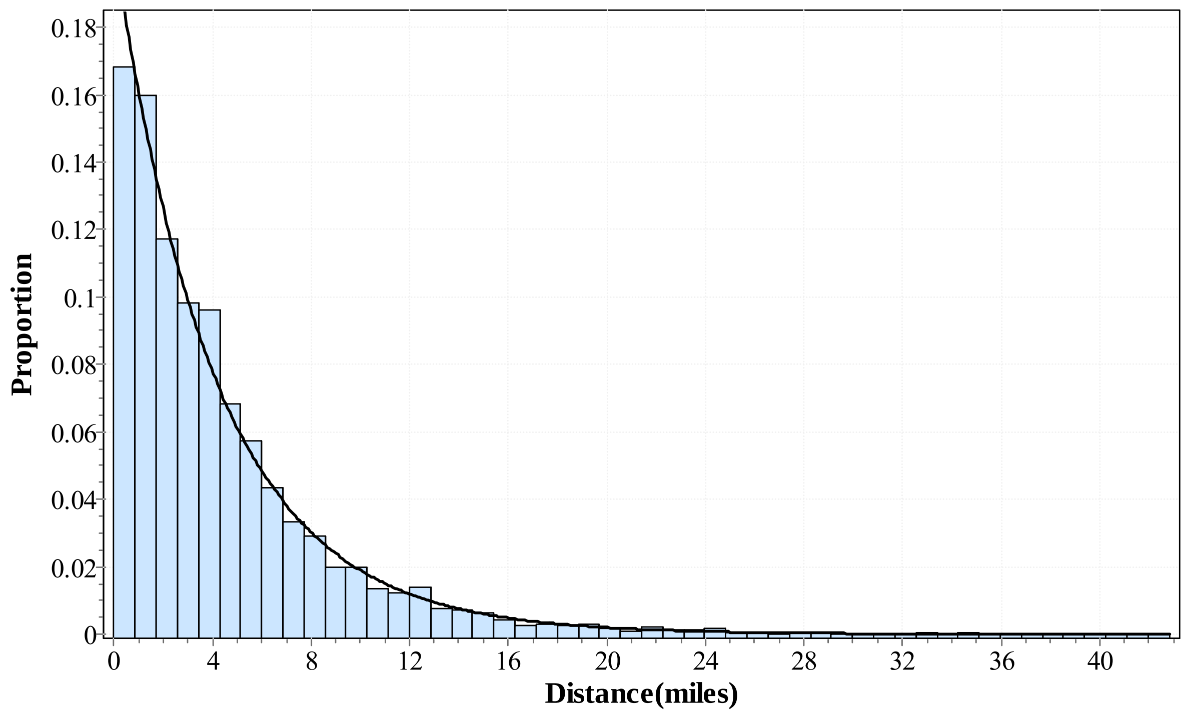


Figure 4.12 Trip Length Distribution of NHBO Trips

Table 4.22 Destination Choice Model of Non Home Based Other Trips

Variable	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**
Constant		0.2654 (0.000)[1]	1.1605 (0.000)	-0.3886 (0.000)
Distance	-0.1250 (0.000)	-0.7368 (0.000)	-0.5121 (0.000)	-0.1765 (0.000)
License or not: 1 is yes, 0 if not		0.5136 (0.000)	0.7005 (0.000)	1.2631 (0.000)
Activity duration		0.0009028 (0.000)	0.001375 (0.000)	0.003413 (0.000)
Retail employment of Destination TAZ		0.000429 (0.000)1	0.0005037 (0.000)	0.0007784 (0.000)
Balance of Destination TAZ	0.6143 (0.000)	0.6143 (0.000)		
Transit coverage rate	0.7856 (0.000)	0.7856 (0.000)		
Road density of Destination TAZ	0.001053 (0.000)	0.001053 (0.000)	0.001053 (0.000)	0.001053 (0.000)
Log(Retail employment of all TAZs in a group)	0.3692 (0.000)	0.3692 (0.000)	0.3692 (0.000)	0.3692 (0.000)
Summary statistics Number of observations =3302 Base model's McFadden Pseudo R squared=0.1426 Final model's McFadden Pseudo R squared=0.1848				

Note: [1]: The value in the parentheses is the p-value of the coefficient.

As is shown by the model results in Table 4.22, NHBO trips' destination choice model includes variables like trip distance, license or not, and activity duration. Trip distance's effect on intrazonal trips is relatively small, and the effect is strongest for alternative 2, then gets weaker as alternatives get further away. These coefficients show that NHBO trips' length distribution should be similar to that of HBW trips, instead of the distribution curve shown in Figure 4.12. This finding reveals that Figure 4.12 is actually not a precise description of the trip length distribution. Both having licenses and longer activity duration make the selection of further TAZs more likely.

In terms of land use variables, a NHBO trip is more likely to travel to a TAZ with more retail employment, and the effect caused by the same amount of retail employment gets stronger when the TAZ gets further away. Balance is significant in alternative 1 and alternative 2, meaning shorter trips are more likely to be induced by higher balance TAZs. Same with the previous models, higher Transit coverage rate help TAZs in alternative 1 and 2 attract more NHBO trips, but has no effect in TAZs more than 2.55 miles away. Higher road density always means more opportunities, thus could attract more trips. Road Density has the same amount of effect in each of the four alternatives.

Table 4.23 Validation Performance of the NHBO Trips' Destination Choice Model

NHBO trips	**Alt. 1**	**Alt. 2**	**Alt. 3**	**Alt. 4**	**Overall**
Actual	123	239	228	221	811
Estimated	74	243	235	259	811
Correctly Estimated	32	129	122	143	426
Prediction Accuracy	26.0%	54.0%	53.5%	64.7%	52.5%

The model validation is shown in Table 4.23. Again with the same reason with other models, the model tends to overestimate the probability of TAZs further away while underestimating the possibility of choosing nearer TAZs.

4.4 Mode Choice Modeling

This section describes the process of building mode choice models to predict travelers' mode choice decisions in the Greater Buffalo-Niagara area. Six transportation modes available in the area are considered as the potential mode choice alternatives, including walk, bike, auto (including both driving and passenger), bus, rail, and taxi/shuttle. In addition to travelers' socio-economic characteristics and trip related attributes, the land use variables of both the origin and destination of a trip are included specifically in order to measure how land use patterns affect travelers' mode choice decisions. Land use variables are grouped into three categories, corresponding to the 3Ds measures of density, diversity, and design indices introduced earlier in Chapter 2.

Considering the important role of intra-zonal trips in the travel demand forecasting and especially when assessing smart growth strategies, mode choice models are developed for intrazonal trips and for interzonal trips separately. The major outcome of this section is two sets of mode choice models for intra- and inter-zonal trips which can be used to predict the probability of a trip choosing a certain mode, given information about the factors affecting the potential mode choice.

4.4.1 Methodology

Mode Choice Set

There are ten modes reported in the 2002 Buffalo-Niagara Regional Household Travel Survey, including: walk, bike, auto driver, auto passenger, metro bus, metro rail, school bus, taxi/shuttle/limousine, motorcycle and others. For the modeling purpose, auto driver and auto passenger are combined as one “auto” mode while “school bus”, “motorcycle” and “others or refuse” are deleted due to either the small market share or few observations. Therefore, six modes are finally considered as the mode choice alternatives, which are walk, bike, auto, bus, rail, and taxi/shuttle. Table 4.24 shows the relationship between the original modes in the raw data and the mode alternatives included in the modeling data as well as the reported market share for each one of them. The basic statistics for the key attributes of these modes, such as the number of observed trips, average trip distances, and trip or activity durations are summarized in Table 4.25.

Table 4.24 Mode Shares from the 2002 Buffalo-Niagara Regional Household Travel Survey

Reported modes	**Combined mode alternatives used in mode choice models**	**Market share (%)**
Walk	Walk	7.5
Bike	Bike	0.53
Auto Driver	Auto	63.55
Auto Passenger		18.42
Metro Bus	Bus	2.96
Metro Rail	Rail	0.44
School bus	Not considered	5.95
Taxi/Shuttle bus/Limousine	Taxi/Shuttle	0.51
Motorcycle	Not considered	0.00
Other and Refuse	Not considered	0.11

Variable summaries of each of the 6 modes are listed in Table 4.25.

Table 4.25 Mode Related Statistics

	Walk	Bike	Auto	Bus	Rail	Taxi/Shuttle bus
Number of trips	1016	73	16173	314	56	50
Average trip distance (miles)	0.88 (1.49)[1]	1.69 (1.32)	5.15 (5.00)	3.87 (3.08)	3.84 (3.26)	2.75 (2.52)
Trip duration (minutes)	9.95 (10.00)	16.40 (22.62)	15.15 (12.56)	34.75 (23.72)	25.20 (23.55)	18.66 (15.08)
Activity duration (minutes)	243.2 (258.9)	173.5 (196.0)	209.6 (236.9)	263.4 (253.8)	183.2 (229.5)	245.7 (283.2)

Note: [1] The number in the parenthesis is the standard derivation.

Nested Structure

In this research, the six mode alternatives are arranged in a choice tree structure to characterize the interrelationships of non-motorized modes (such as walk and bike), motorized modes (i.e., auto, transit, and taxi/shuttle bus) and transit modes (i.e., rail and bus). Three choice trees, as shown in Figure 4.13, Figure 4.14, and Figure 4.15, were tested and the best one was selected, in terms of higher R squared and IV parameters. The results showed that the model of choice tree 3 had the highest R squared. In terms of IV parameters, the model of choice tree 1 has an IV parameter as 4, far beyond the reasonable range, and the model of choice tree 2 has two IV parameters more than 1, while the model of choice tree 3 had only one of the four IV parameters slightly out of range (IV parameter of motorized branch is 1.035). Thus choice tree 3 is selected as the best tree structure.

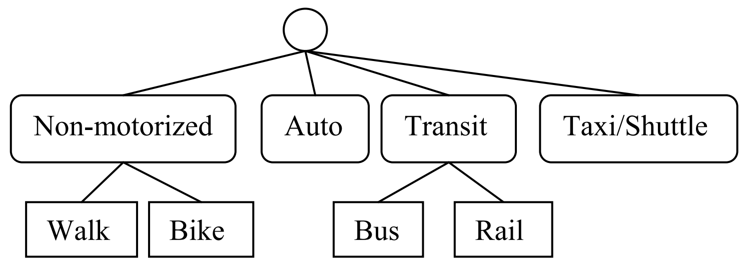


Figure 4.13 Nest Tree Structure 1

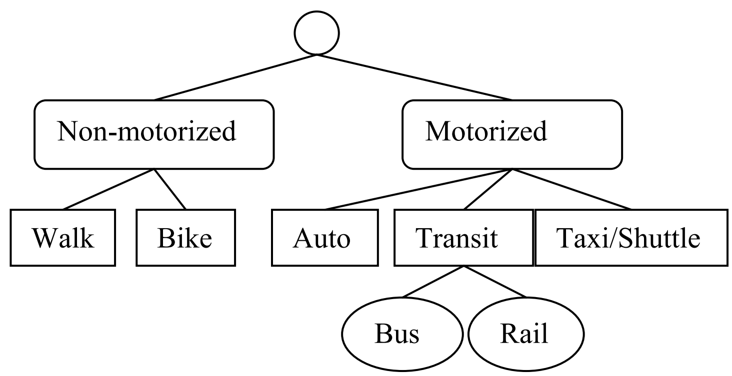


Figure 4.14 Nest Tree Structure 2

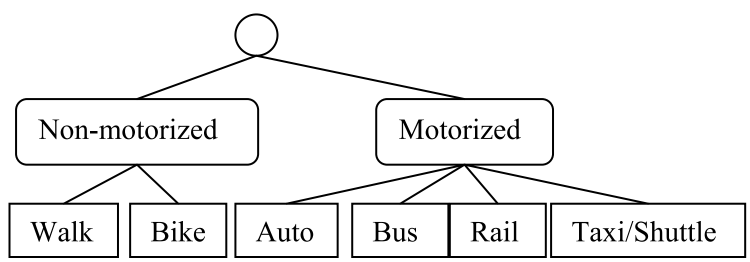


Figure 4.15 Nest Tree Structure 3

Why Not Use Purpose-Specific Models

Both the trip generation and destination choice are purpose-specific models. Should mode choice model also be purpose specific? A data pre-analysis was performed to answer this question. The mode share of each purpose is shown in TABLE 4.26. It shows that the mode market share does not vary much among different trip purposes (especially when the auto driver and the auto passenger modes are combined).

TABLE 4.26 Mode Share Percentage of Each Trip Purpose

Mode	Mode	HBW	HBShop	HBSRO	NHBW	NHBO
1	Walk	2.8	5.7	8.6	7.5	6.4
2	Bicycle	0.4	0.6	0.7	0.3	0.2
3	Auto Driver	85.6	68.6	63.2	81.6	68.6
4	Auto Passenger	6.6	20.0	23.6	6.2	18.3
5	Metro bus	4.3	3.8	1.9	3.1	3.0
6	Metro Rail	0.3	0.0	0.2	0.5	1.1
7	School bus	0.0	0.0	1.2	0.2	1.7
8	Taxi/Shuttle	0.0	1.1	0.4	0.5	0.5
9	Motorcycle	0.0	0.0	0.0	0.0	0.0
10	Other/Refuse	0.0	0.1	0.2	0.1	0.1

As the selection of non-motorized modes is highly dependent on travel distance, we analyzed the mode share of intrazonal trips and interzonal trips, as shown in Table 4.27. The walk market share in intrazonal trips is 28.9%, while only 3.5% in interzonal trips. The market share of Auto Driver, on the other hand, is much higher in inter-zonal trips than that in intra-zonal trips.

Table 4.27 Mode Share Percentage of Intrazonal and Interzonal Trips

ID	Mode	Intrazonal	Interzonal
1	Walk	28.9	3.5
2	Bicycle	1.2	0.4
3	Auto Driver	47.6	66.5
4	Auto Passenger	16.2	18.8
5	Metro bus	0.5	3.4
6	Metro Rail	0.3	0.5
7	School bus	4.8	6.2
8	Taxi/Shuttle bus/Limousine	0.3	0.5
9	Motorcycle	0.0	0.0
10	Other and Refuse	0.2	0.1

Given the discussions above, we abandoned building purpose-specific models, but instead we develop two models: one for intra-zonal trips and the other for inter-zonal trips.

4.4.2 Data Source and Data Assembly

Data Assembly

The steps to generate the dataset are as follows:

1) Generate the TAZ-based land use attributes from the parcel land use map.
2) The two trip ends, which are coded in "location", are projected onto the TransCAD TAZ and traffic network map. The shortest network distance for each trip is generated using a GIS tool. As there are multiple "locations" in one TAZ, we also get the trip distance of the intrazonal trips.
3) We find the TAZ for each trip end's location. Following this, the trips recorded in the travel survey are categorized into two groups: intra-zonal trips and inter-zonal trips, and each group

is stored in a separate file. Specifically, 1986 intra-zonal trip observations and 2665 inter-zonal trip observations are randomly sampled from the two data files to build the models.

4) Each trip is duplicated six times, with each record having a separate mode. Given the travel time of the chosen mode, the travel times of the other five modes are calculated using the speed ratios.
5) The trip datasets of both intra-zonal and inter-zonal trips are matched with the person/household socio-demographic file and TAZ level land use file to get the socio-demographic characteristics, and land use attributes for each of the trip ends.

The final intra-zonal trip dataset thus includes 1986 trip observations (1986*6 = 11916 records), whereas the inter-zonal trip dataset includes 2665 trip observations (2665*4=15990 records). Around 80% of the observations in each dataset are used to calibrate the model, and the remaining 20% are used for model validation.

Estimating the Trips' Travel Time

As an indicator of travel cost, travel time is a very important variable influencing travelers' mode choices. Each trip observation in the 2002 Buffalo-Niagara Regional Household Travel Survey has a chosen mode and the travel time for that mode. However, the travel time for the alternative modes needed to be estimated for the purposes of model building. Considering the limitation of data source, we use a simplified method to generate the travel times. Specifically, we use the survey data to define the average speed ratios of the six modes. Then given the actual travel time of the chosen mode, we calculate the travel time of all the other five modes. The process of generating the speed ratios is shown in Table 4.28. We split the survey trips into all the six modes, and calculate the average speed of the trips in each mode. Then based on the calculated speeds, we set some reasonable values, as shown in the fourth row of Table 4.28. The set of travel time ratios is: walk: bike: auto: bus: rail: taxi/shuttle = 30 : 10 : 3 : 10 : 6 : 10.

Table 4.28 Travel Time Ratios of the Six Modes

Mode	**Walk**	**Bike**	**Auto**	**Bus**	**Rail**	**Taxi**
Average speed/mph	2.68	11.76	23.13	9.77	15.46	9.48
Standard Derivation	1.39	16.48	25.31	16.39	16.47	7.4
Assumed speed/mph	3	9	30	9	15	9
Travel time ratio	30	10	3	10	6	10

4.4.3 Model Formulation

The nested logit model assumes that multiple choices share unobserved attributes, and wraps these alternatives into a nest. The utility of an elemental alternative m in nest d is expressed as follows.

$$\boldsymbol{U_{dm}} = \overline{\mathbf{V}_{\boldsymbol{d}}} + \overline{\mathbf{V}_{\boldsymbol{m}}} + \overline{\mathbf{V}_{\boldsymbol{dm}}} + \overline{\boldsymbol{\varepsilon}_{\boldsymbol{m}}} + \overline{\boldsymbol{\varepsilon}_{\boldsymbol{dm}}}$$

Where

$\overline{V_d}$ = The systematic component of utility common to all elements using nest d.

$\overline{V_m}$ =The systematic component of utility common to all elements using alternative m.

$\overline{V_{dm}}$ = The remaining systematic component of utility specific to the combination *(d,m)*.

$\overline{\varepsilon_m}$ = The unobserved components of the total utility attributable to the alternative m.

$\overline{\varepsilon_{dm}}$= The random utility component.

Several assumptions are further made as listed below:

1) $\overline{\boldsymbol{\varepsilon_m}}$ and $\overline{\varepsilon_{\mathrm{dm}}}$ are independent for all d∈ $D_n$and m∈ M_n**.**
2) The terms $\overline{\varepsilon_{dm}}$ are independent and identically Gumbel distributed with scale parameter μ^d.
3) $\overline{\boldsymbol{\varepsilon_m}}$ is distributed to that $max_{\mathrm{d}\in D_{nm}} U_{dm}$ is Gumbel distributed with scale parameter μ^m.

With these assumptions, the probability of choosing an elemental alternative m in nest d can be represented by a nested logit model as shown below:

$$P_n(d) = \frac{e^{\mu^d(\overline{V_d}+V'_d)}}{\sum_{d' \epsilon D_n} e^{\mu^d(\overline{V_{d'}}+V'_{d'})}}$$

$$\boldsymbol{P_n(m|d)} = \frac{e^{\mu^m(\overline{V_m}+\overline{V_{dm}})}}{\sum_{m' \epsilon M_{nd}} e^{\mu^m(\overline{V_{m'}}+\overline{V_{dm'}})}}$$

Where:

$P_n(m|d)$=The conditional probability of elemental alternative m being chosen given the choice set M_{nd} for observation n;

$P_n(d)$= The conditional probability of nest d being chosen given the choice set D_n for observation n;

$\overline{V_d}$ = The systematic component of utility common to all elements using nest d.

$\overline{V_m}$ =The systematic component of utility common to all elements using alternative m.

$\overline{V_{dm}}$ = The remaining systematic component of utility specific to the combination (d,m).

$$V'_d = \frac{1}{\mu^m} ln\, (\sum\nolimits_{m \epsilon M_{nd}} e^{\mu^m(\overline{V_m}+\overline{V_{dm}})})$$

$\frac{\mu^d}{\mu^m} = \sqrt{1 - coor(U_{dm}, U_{dm'})}$ should be within 0 and 1.

The nest structure could have more than two levels. More information about the nest logit model can be found in Ben-Akiva's book (Ben-Akiva and Lerman 1985).

4.4.4 Mode Choice Model Results

Mode Choice Model of Intrazonal Trips

Table 4.29 Mode Choice Model of Intrazonal Trips

Variable	Non-motorized		Motorized			
	Walk	Bike	Auto	Bus	Rail	Taxi/ Shuttle
Alternative specific constant	5.2785 (0.000)[1]	-0.3691 (0.000)		-3.7438 (0.000)	-4.1742 (0.000)	-504.65 (0.000)
Travel time	-0.2610 (0.000)	-0.2610 (0.000)	-0.2610 (0.000)	-0.1068 (0.000)	-0.1068 (0.000)	-0.4138 (0.000)
Number of vehicles owned per person	-2.8792 (0.000)	-2.8792 (0.000)		-7.8823 (0.000)	-7.8823 (0.000)	-4.8946 (0.000)
Traveler's age	0.00348 (0.000)					5.1593 (0.000)
Average personal income in the household (1000 dollars)		0.01722 (0.000)				
Activity Duration						0.002614 (0.000)
TAZ population density	0.864E-4 (0.000)		-0.526E-4 (0.000)			
TAZ employment density	0.107E-4 (0.000)		-0.367E-4 (0.000)			
TAZ dissimilarity index			-1.1521 (0.000)			
TAZ employment entropy	2.7903 (0.000)					
Existence of ramp to freeway (1: yes 0: no)			0.1558 (0.000)			
Transit coverage				0.7798	0.7798	

				(0.000)	(0.000)	
IV parameter	0.9447 (0.000)		1.0348 (0.000)			
Summary statistics Number of observations =1600 Base Model McFadden Pseudo R squared=0.8731 Final Model McFadden Pseudo R squared=0.8800						

Note: [1] All the variables are significant at 0.01 confidence level.

The values and signs of the estimated coefficients provide rich implications about the impacts of the different factors on mode choice. In the intra-zonal trips' mode choice model (Table 4.29), the estimated coefficients of trip travel time in all the six utility functions are negative, meaning longer travel time decreases the likelihood of selecting a mode. More specifically, transit users are not as sensitive to travel time increase as users of non-motorized modes and auto, indicating transit users have lower value of time. On the other hand, taxi users have higher value of time.

Higher auto ownership results in less usage of all modes except auto, and especially less usage of transit. This is because transit is a substitute for auto, thus people without cars become dependent on transit. Different from transit, walk and bike are a supplement, not a substitute to auto, as they are available for only short trips. What's more, considering the positive impact of average personal income on bike usage, it appears that biking is regarded as a means for exercise and healthy living among the middle- and upper- class residents. Age seems to be positively related to walk and taxi, maybe senior citizens walk near their house for leisure, and use taxi more for longer trips. Longer activity duration actually makes taxi more attractive.

In terms of land use variables, as can be seen, TAZ density and diversity do change travelers' mode choices. Higher density TAZs encourage walk and discourage auto. Higher dissimilarity index, or more diversified TAZs discourage auto usage, and higher employment entropy encourage walking. The existence of a ramp in a TAZ is related to automobile travel. This is evidence that highway facilities induce more auto-oriented long distance trips, or that easy access to highway could attract people who like driving. Finally, it is easy to understand that accessibility to transit stops (higher parameter of Transit coverage) would encourage usage of bus and rail.

Table 4.30 Validation Performance of the Intrazonal Trips' Mode Choice Model

Intrazonal trips	**Walk**	**Bike**	**Auto**	**Bus**	**Rail**	**Taxi**	**Overall**
Actual	98	3	266	1	17	1	386
Estimated	115	0	268	0	1	2	386
Correctly estimated	85	0	249	0	0	0	334
Percentage	86.7%	0	93.6%	0	0	0	86.5%

The developed mode choice model of intra-zonal trips is validated using data from the same survey. As can be seen from Table 4.30, 86.5% of the trip observations for validation are correctly estimated.

Mode Choice Model of Inter-zonal Trips

Table 4.31 Mode Choice Model of Interzonal Trips

Variable	Non-motorized		Motorized			
	Walk	Bike	Auto	Bus	Rail	Taxi/ Shuttle
Alternative specific constant	6.7287 (0.000)[1]	1.9793 (0.000)		0.1277 (0.000)	-5.6021 (0.000)	-0.9257 (0.000)
Travel time	-0.1797 (0.000)	-0.1797 (0.000)	-0.1385 (0.000)	-0.0544 (0.000)	-0.0544 (0.000)	-0.1321 (0.000)
Number of vehicles owned per person	-2.5176 (0.000)	-4.2726 (0.000)		-9.0987 (0.000)	-5.0475 (0.000)	-6.4991 (0.000)
Female or not	-0.1616 (0.000)	-2.4571 (0.000)				0.5744 (0.000)
Age	-0.0105 (0.000)					0.0541 (0.000)
Number of trips made per person per day			0.1639 (0.000)			
DTAZ population density	0.1664E-4 (0.000)	0.8729E-4 (0.000)				
DTAZ employment density	0.1664E-4 (0.000)					
OTAZ employment entropy			0.6107 (0.000)			
Existence of ramp to freeway in OTAZ (1: yes 0: no)			0.7200 (0.000)			
DTAZ Transit Coverage				3.0520 (0.000)	4.6174 (0.000)	
IV parameter	0.8867 (0.000)		0.9228 (0.000)			

Summary statistics
Number of observations =2132
Base Model McFadden Pseudo R squared=0.9168
Land Use Model McFadden Pseudo R squared=0.9222

Note: [1] All the variables are significant at 0.01 confidence level.

A separate model choice model is identified for inter-zonal trips (Table 4.31). The trip travel time parameters are all negative in all the six utility functions. Different from the intra-zonal trips' mode choice model, non-motorized modes are most strongly influenced by the increase of travel time. This might be because inter-zonal trips could have distances as long as 40 miles and travel time as long as one hour. With the increase of trip distance and travel time, the probability of choosing non-motorized modes drops down to zero. Same with intra-zonal trips' model, transit users have the lowest value of time.

The effect of vehicle ownership on inter-zonal trips' mode choice is very similar to that on intra-zonal trips. The only difference is that the parameter of auto ownership in bus is larger (in terms of absolute values) than that in rail. This might be because the service coverage area for buses in Buffalo is much larger than rail, and as a result, it is more likely to be taken as auto's substitute (there is only one rail line in this area, which mainly serves downtown Buffalo).

Males are more likely than female travelers to walk or bike. On the other hand, females tend to use taxi/shuttle bus more than males. Senior citizens take the taxi more and walk less, while in intra-zonal trips' mode choice model, senior citizens walked more. This may be because shorter intra-zonal trips may be primarily for leisure and exercise for senior citizens, while longer inter-zonal trips are not. People who need to make multiple trips per day usually prefer the auto, since it provides better flexibility.

In terms of land use variables, both the origin TAZ's and destination TAZ's land use variables influence travelers' mode choices. Higher population density and/or higher employment density in destination TAZ promote walk. Furthermore, population density of destination TAZ is positively related to bike usage. If origin TAZ's employment entropy is higher, travelers starting from that zone actually drive more and walk less; no obvious reason was readily apparent to explain this. Existence of ramp to freeway encourages automobile usage, and transit coverage of a TAZ makes market share of bus and rail higher.

Table 4.32 Validation Performance of the Interzonal Trips' Mode Choice Model

Interzonal trips	**Walk**	**Bike**	**Auto**	**Bus**	**Rail**	**Taxi**	**Overall**
Actual	17	2	503	7	3	1	533
Estimated	15	0	511	5	0	2	533
Correctly predicted	12	0	500	3	0	1	516
Percentage	70.6%	0	99.4%	42.9%	0	100%	96.8%

The mode choice model of inter-zonal trips was validated using data from the same survey (see Table 4.32). As can be seen, 96.8% of the 533 observations are correctly estimated.

4.5 Case Study

As a demonstration of how the enhanced travel demand forecasting method developed in this research may be applied, the method was used to evaluate several hypothetical smart growth land use scenarios. Specifically, the method was first applied to the base case Buffalo model. Following this, several smart growth strategies were assumed, and their likely impact was evaluated. By comparing the scenario analysis results, insight are gleaned regarding the impact of smart growth strategies on travel demand and the applicability of the proposed method. The details are shared as below.

4.5.1 Enhanced Travel Demand Forecasting Framework and Steps

The developed enhanced travel demand forecasting modeling framework to evaluate smart growth strategies is shown in Figure 4.16. As can be seen, it follows the modeling structure of the traditional four-step travel demand models, but is improved by taking into consideration individual travelers' travel behaviors, particularly in the steps of destination choice and mode choice, as described above. When applied to evaluate a land use scenario, the framework starts with the data collection in order to get the travel and land use related information. Then, trip generation models are run to predict the number of trips produced by each TAZ for each trip purpose. After the total trips are predicted, a trip population synthesizer is used to generate the profile for each trip in terms of the socio-economic characteristics of the household and the person who make the trip and the trip-related attributes, such as activity duration. In addition, the land use attributes for the origin and all the potential destinations of each trip are added to the synthesized trip data. Later on, destination choice models are used to predict the probability for each synthesized trip to choose a destination TAZ in the area. Furthermore, the mode choice models will be run to estimate the mode choice decision for each trip. These individual-trip-based destination choice and mode choice decisions are then aggregated to get the OD trip tables by trip purposes by modes. As the last step, the OD demand is assigned to the transportation network to obtain the performance indicators such as VMT and VHT. The key steps of the modeling framework are discussed in detail below. All the procedures are implemented using a Java code.

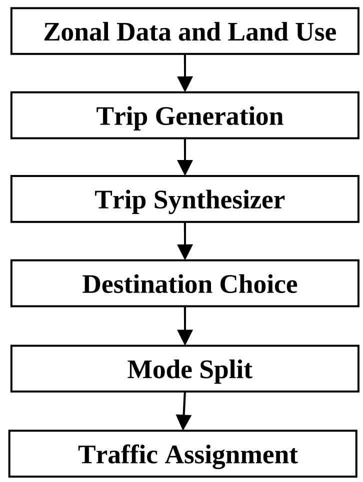


Figure 4.16 Enhance Travel Demand Forecasting Framework

Trip Generation

As the first modeling step in the framework, the trip generation models developed in section 4.2 are used to predict the total number of trips produced (called trip production) from each TAZ for each of six trip purposes. The estimated trip productions are shown in Table 4.33 with 10 TAZs as the examples. In total, 2,676,784 trips are produced by Erie County during a typical weekday. These estimated trip productions will remain the same for all the land use scenarios.

Table 4.33 Estimated Trip Productions of TAZs during a Typical Weekday (Trips)

TAZ	Trip purposes					
	HBW	**HBShop**	**HBSR**	**HBO**	**NHBW**	**NHBO**
1	302	551	359	334	753	2473
2	16	431	213	0	521	1153
3	16	428	211	0	521	1156
4	16	431	214	0	430	686
…						
161	1305	647	2724	150	948	161
162	586	444	1251	145	394	162
163	537	548	1225	268	809	163
164	345	516	830	468	1622	164
165	357	763	1098	783	3528	165
166	1811	823	3797	138	824	166
……						
Sum	464479	293800	303838	730955	142860	740830

Trip Synthesizer

After the trip production from each TAZ is estimated, the trip synthesizer is used to generate the profile for each trip in terms of the trip maker's socio-economic characteristics, the associated household's attributes and trip related attributes. In terms of trip makers' and households' attributes, the trip synthesizer generates 14 socio-economic variables for each trip as shown in Table 4.34, including the trip maker's age, gender, whether a student or not, number of jobs, household size, household income, car ownership, and so on. In addition, activity duration is generated for each trip as the main trip attributes. When determining the specific value for each attribute, the trip synthesizer does random sampling from the survey data. Specifically, what it does is to randomly select a value from the observed distribution of an attribute and then assign it to a target trip. By doing so for each attribute of each trip, we generate a trip profile data as the outcome of trip synthesis.

It needs to note that this random synthesis process could generate unrealistic trip attributes. For example, when independently assigning values to HWORK (number of workers in a household), HSTUD (number of students in a household), and HHSIZ (household size) for a trip, we may have values assigned to HWORK and HSTUD larger than HHSIZ. To avoid these inconsistent assignments, we added constraints to address the relationships between different attributes. For instance, we added a constraint in the trip synthesizer to make sure that HWORK and HSTUD are always less than or equal to HHSIZ. We did so for other attributes as well.

To give more details about how the trip synthesizer works, we use the HBW trips of TAZ 25 as an example. As the initial input, 2,508 HBW trips are predicted to be produced from TAZ 25 as the result of the trip generation step. When using the trip synthesizer to generate the profile for each of the 2,508 HBW trips, we firstly obtained all the five HBW trips originated at TAZ 25 and their associated information from the 2002 Buffalo-Niagara Regional Household Travel Survey (Table 4.34). Then, we used the five trips as the seeds to determine the attributes for each of the 2,508 HBW trips. For example, we randomly selected a value (which is 3) from the five reported HHSIZ values and assigned it to HHSIZ of the first predicted trip, and did the same thing for other attributes of the trip as well as for other trips. By doing so, we synthesized the profiles for the 2,508 trips as shown in Table 4.35.

Table 4.34 Five HBW Trips Originated at TAZ 25 from the Survey

TAZ	HHSIZ	HHVEH	HWORK	HSTUD	…	PERINCM	FEMALE	AGE	JOBS	STUDE	NTRIPS	ACTDUR
25	3	3	3	1		19.8	0	45	2	2	1	622
25	3	3	3	1		19.8	1	45	1	2	3	410
25	1	0	1	0		20	1	60	1	2	2	570
25	2	0	1	0		17.5	1	46	1	2	6	365
25	2	0	1	0		17.5	1	46	1	2	6	65

Table 4.35 Synthesized HBW Trip Set Originated at TAZ 25

TAZ	HHSIZ	HHVEH	HWORK	HSTUD	…	PERINCM	FEMALE	AGE	JOBS	STUDE	NTRIPS	ACTDUR
25	3	0	3	1		19.83	1	46	1	2	6	410
25	2	0	1	0		17.50	0	46	1	2	3	622
25	2	3	1	1		19.83	1	46	1	2	2	570
25	1	3	1	0		17.50	1	46	1	2	6	65
25	2	0	1	1		19.83	1	46	2	2	1	622
25	2	3	1	0		19.83	0	60	1	2	2	365
25	1	0	1	1		20.00	1	60	1	2	6	410
25	3	3	3	0		19.83	1	60	1	2	3	570
25	2	0	1	0		20.00	1	60	1	2	6	65
…												

Destination Choice

This step is to estimate the probability for a trip to choose a destination TAZ given a land use scenario. Theoretically speaking, each trip can go to any of the 402 TAZs in the Erie County as the destination. Therefore, the choice probability should be estimated for each TAZ for a synthesized trip. However, to do so is very computationally challenging due to the large number of trips involved (i.e., 2,676,784 synthesized trips) and the large destination alternative set (i.e., 402 TAZs) available to each trip. To solve this issue, we used the stratified random sampling techniques, as we did for the calibration of destination choice models, to select four destination TAZs as the representative choice alternatives for each trip. Figure 4.17 shows the four distance-based strata we used to sample destination alternatives, and the stratification strategy is the same as the one we used to calibrate destination choice models. After sampling four alternative destination TAZs for each trip, the land use attributes for these TAZs were retrieved and attached to the trip profile data estimated from the trip synthesis step. By doing so, we added land use attributes, particularly for alternative destination TAZs, to the original trip profile data.

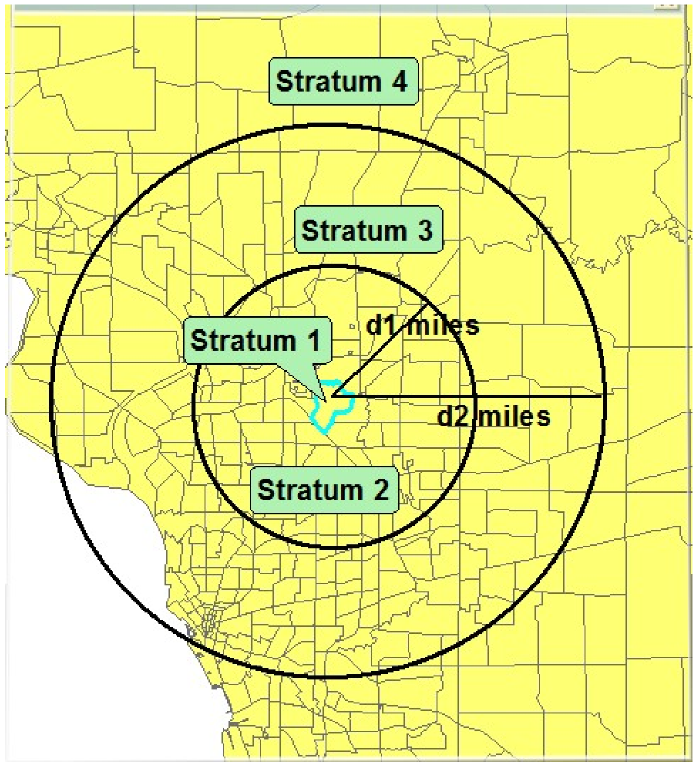


Figure 4.17 The Distance-Base Stratified Sampling Technique

After the data preparation, the destination choice models were run for each trip of a given trip purpose, to estimate the probability of that trip choose a destination TAZ.

Mode Choice

After running the destination choice models to derive the probability of a trip choosing a destination, the mode choice models were run to determine the probability of that trip choosing one of the six available transportation modes. Specifically, for each trip, the intra-zonal mode choice model was applied to the first sampled destination TAZ that is the same as the origin of the trip, and the inter-zonal mode choice model was used to other destination TAZs. The direct outcome of this step is the estimated probability for a trip to visit a destination TAZ by using certain mode. These trip-based choice probabilities were then aggregated to get the OD trip tables by modes.

It needs to mention that trip travel times are a key input used to estimate the mode choice probabilities. For inter-zonal trips that involve the destination TAZs other than the origin TAZ of a trip, we used the TransCAD multiple shortest path module to generate trip travel times. For intra-zonal trips, it was a bit challenging since TransCAD assumes zero travel time for any intrazonal travel. To solve this problem, we designed a random distance generator to generate the trip distance for each intra-zonal trip, assuming the distance could be anywhere from zero to the diameter of the origin TAZ. Then we divided each trip distance by the average speed of each mode to get the travel times by modes for each trip. Several random distance generators were tested, the best one that returns the most consistent intra-zonal mode market share was used at the end.

Traffic Assignment

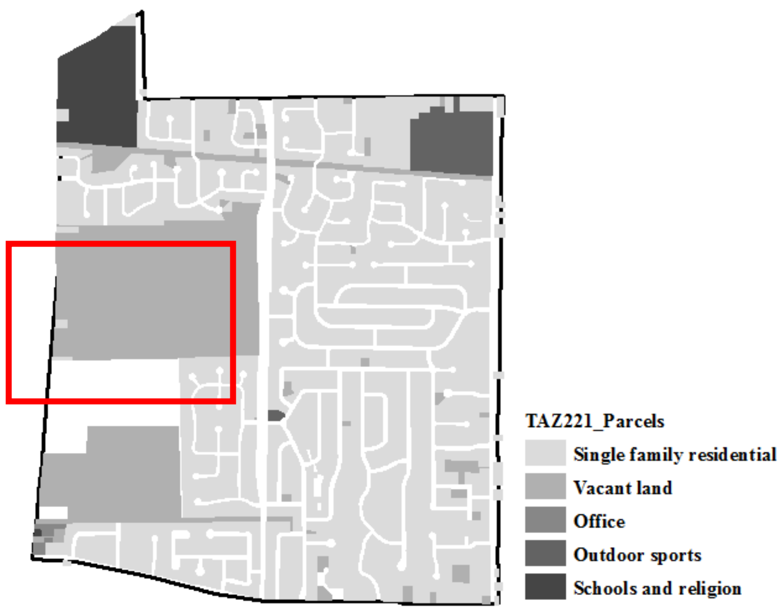
As the last step of the travel demand forecasting process, for each scenario, we used the all-or-nothing assignment method to load the estimated auto trips to the roadway network in the region. Bus and rail trips were not assigned as they contribute very little to the major performance measures such as VMT and VHT. Trips of taxi/shuttle bus were also not assigned either due to their negligible market shares.

4.5.2 Scenario Analyses

Three smart growth land use scenarios are designed to test the feasibility of the proposed travel demand forecasting approach. The first scenario assumes that a shopping center would be added to a suburb residential area. In the second scenario, a mixed land use smart growth policy will be promoted all across Erie County, and as a result the land use dissimilarity index in the whole county improves. The last scenario focuses on transit-oriented smart growth strategies, assuming the implementation of these strategies will increase the transit coverage all across the whole county. The detailed analyses are discussed below.

Smart Growth Scenario One: Redevelopment in a TAZ

This scenario assumes there would be a major redevelopment in TAZ 221. TAZ 221 is located in the north eastern part of Erie county, and is a suburban area dominated by residential land uses. It is chosen as a representative of single-land-use zones that are typical in Erie County.



Note: The red rectangle means the shopping center.

Figure 4.18 Parcel Level Map of TAZ 221

As shown in Figure 4.18, TAZ 221 is composed of single family dwelling units, very limited office space, an outdoor sports area, and two vacant lands. In this scenario, we assume a shopping plaza, that involves multiple types of employment opportunities such as retail and wholesale, will be built in the vacant land. Due to the addition of the shopping center, the demographics and land use variables of this TAZ are changed as shown in Table 4.36.

Table 4.36 Changes of Land Use Variables in Scenario One

Demographics and Land Use Attributes	Base Case	After Building the Shopping Center
Retail employees	0	400
Wholesale employees	16	66
Other employees	597	795
Total employees	648	1296
Employment density (employees/squared miles)	290.58	581.16
Normalized employment to population ratio	-0.7845	-0.611
Employment entropy	0.3035	0.658
Retail land use (squared feet)	0	500000
Land use entropy	0.2979	0.6
Non-work land use entropy	0.2857	0.55
Dissimilarity index	0.1552	0.5

Then we run the model. The travel behavior changes after the redevelopment are shown in Table 4.37 and Table 4.38. The number of intrazonal trips in TAZ 221 almost doubles, which means the travelers in TAZ 221 are more likely to make shorter interzonal trips and intrazonal trips in the scenario. As a result, VMT of the trips produced by TAZ 221 decreases by 15.1%, and VHT decreases by 4.63%. Table 4.38 shows that the market share of the different modes does not change much. Thus, it seems that that particular scenario’s land use changes influenced travel behavior mainly by changing destination choice instead of mode choice.

Table 4.37 Overall Travel Demand Changes (Scenario one)

	Base case	**Scenario one**	**Percent Change**
Trip production	20298	20298	0
Intrazonal trips	1346	2627	**95.2**
VMT (miles)	101381	86047	**-15.1**
VHT (hours)	2850	2718	**-4.63**

Table 4.38 Mode Shifts (Scenario One)

Modes	**Base case**		**Scenario 1**	
	Market share (%)	**% among intrazonal trips**	**Market share (%)**	**% among intrazonal trips**
Walk	2	2	2	2
Bike	3	17	4	18
Auto	92	81	92	79
Bus	3	0	2	0
Rail	0	0	0	0
Taxi	0	0	0	0

Smart Growth Scenario Two: Mixed Land Use Policy

This scenario assumes a mixed land use policy is promoted in the whole Erie county area. This policy aims to change the isolation of land use function, and makes different land uses more mixed together in a finer grain. It is also assumed that the proportion of employment type and proportion of each land use' are do not change, so that only the dissimilarity index increases while employment entropy and land use entropy remain the same. The enhanced travel demand forecasting approach is run five times, with the dissimilarity index steadily increasing. The VMT reduction of each trip purpose and overall reduction are shown in Figure 4.19.

As is shown in Figure 4.19, VMT of HBW and NHBW trips are most strongly influenced by land use diversification strategies. A 10% inflation of dissimilarity index leads to 0.3% reduction of HBW trips' VMT, and 0.42% reduction of NHBW trips' VMT. The relationship between VMT reduction and dissimilarity index inflation is approximately linear. And the elasticity of total VMT reduction to dissimilarity index inflation is -0.01. Compared with the elasticities given by INDEX-4D method (Criteron_Planner&Engineers and Fehr&Peers_Associates 2001), where the elasticity of daily VMT with respect to a diversity indicator is 0.05, this result is a much lower value. This might be a result of the sparse land uses in this area and people's dependence on private vehicles.

There is one problem within this scenario. It is very rare in real life when only dissimilarity index changes while all other variables remain the unchanged. So this scenario is an ideal simplification of real life. Thus more complex scenario designs, which consider the interrelations among the variables, are needed in future work.

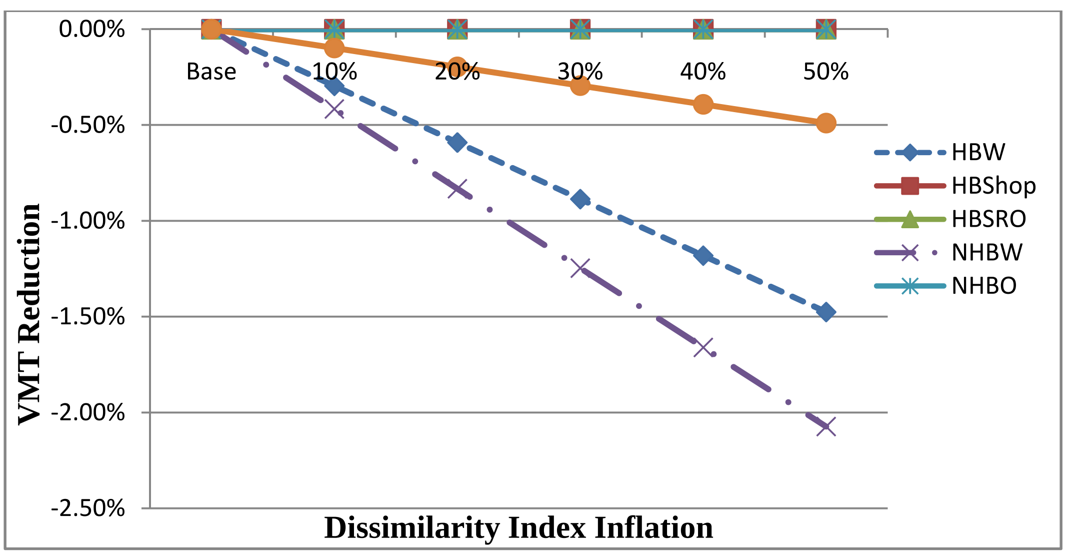


Figure 4.19 VMT Reduction in Scenario Two

Smart Growth Scenario Three: Transit-oriented Policy

Another important set of smart growth strategies aim to build transit-oriented neighborhoods to attract more travelers to bus or rail. This scenario assumes there is a transit-oriented policy implemented in the whole Erie county, so that transit coverage rate, which denotes the proportion of a TAZ within 0.25 miles from transit stops, increases in all the 402 TAZs. Same with scenario two, the model is run five times, with each transit coverage rate has an increasing inflation. The VMT reductions are shown in Figure 4.20.

As is shown in Figure 4.20, VMT of HBShop trips are mostly influenced by transit-oriented developments. The elasticity of HBShop trips' VMT reduction to transit coverage rate inflation is -0.1, which is relatively high. VMT of HBSRO trips is also reduced when transit coverage rate increases, but the reduction is not as big as the VMT reduction of HBShop trips. The elasticity of total VMT to transit coverage rate is -0.015. In both the two trip purposes, the relationship between VMT reduction and transit coverage rate inflation is approximately linear.

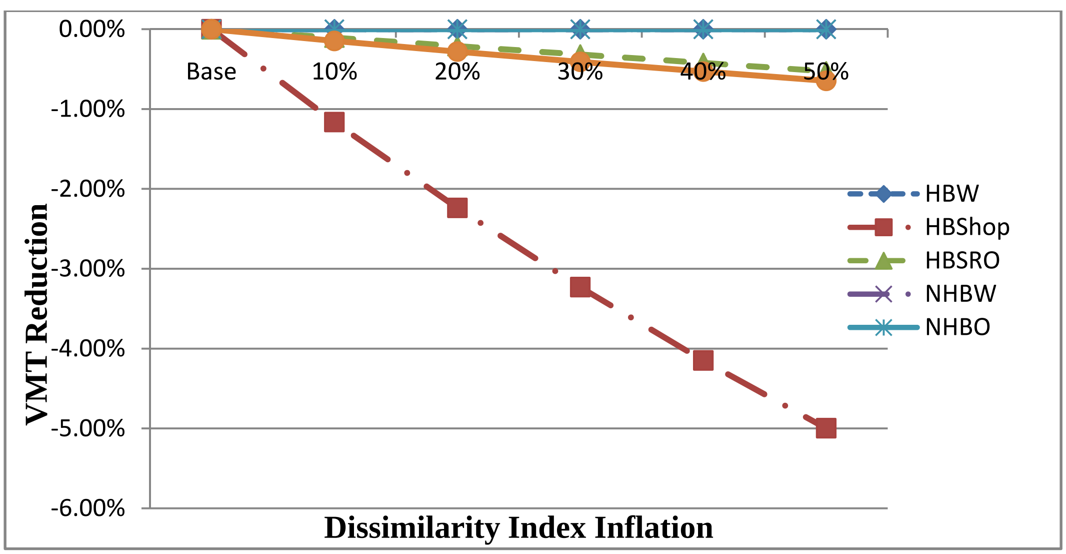


Figure 4.20 VMT Reduction in Scenario Three

4.5.3 Summary of Scenario Analyses

The three scenario studied show that travel behavior, which is denoted by VMT and VHT, is impacted by different land use patterns. To be more specific, redevelopments in a low density residential neighborhood induce more intra-zonal trips in the TAZ, and reduce the daily VMT of this TAZ by 15.1%. Higher dissimilarity land uses in Erie county can shorten the HBW and NHBW trips by increasing the probability of choosing destination TAZs less than 4.65 miles. Provision of transit infrastructure to Erie county impacts the destination choices of HBShop and HBSRO trips. Higher transit coverage rate is associated with shorter HBShop and HBSRO trips. Both higher mixed land uses and transit-oriented land uses could reduce the overall VMT of Erie county.

The sensitivity of this model to smart growth strategies is moderate at most. The elasticity between VMT and dissimilarity index is -0.01, and the elasticity between VMT and transit coverage rate is -0.015. Compared to the elasticities in INDEX-4Ds method (Criteron_Planner&Engineers and Fehr&Peers_Associates 2001) where the elasticity between VMT and 4Ds variables is at least 0.04, the sensitivity of the enhanced travel demand forecasting approach is smaller. Still, the scenario analysis proves that the enhanced travel demand forecasting method could be used to evaluate the impacts of smart growth strategies. Considering the fact that most MPOs in US are still using traditional four-step method, this finding means they can improve their model with relatively small consumption of time and cost to make it sensitive the smart growth strategies, instead of spending a lot of money in reconstructing activity-based model.

5. CONCLUSIONS AND RECOMMENDATIONS

This section briefly summarizes the main conclusions derived from this stud. The section is divided into two parts. The first part lists the conclusions derived from the study's work involving the development of the post-processor method for assessing the likely impacts of smart growth on travel behavior. The second part, on the other hand, focuses on the enhanced four-step travel demand forecasting process developed.

5.1 Post-processor Method Conclusions

In this study, a method of relating zonal travel behavior to built environment factors was presented and applied to a Buffalo, NY study area. Travel behavior was quantified in three ways: mode choice, daily VHT per household, and daily VMT per household. The built environment was characterized by iterative testing and refinement of variables based on a variety of data sources. Linear regression was used to create models relating variations in travel behavior to the built environment.

The resulting models may be applied to transportation planning as tools to estimate the potential impact of a development plan on travel behavior. Changes in TAZ-level mode choice, VMT and VHT may be estimated, provided the development plan is expected to significantly alter the built environment in a way that can be quantified in the explanatory variables. The models themselves may be examined in the absence of a development plan to serve as guidance in future planning. Among the main conclusions of the study are:

1. Zonal mode choice is highly correlated to built environment factors, especially those related to density or design, even when controlling for relevant demographics such as household vehicle ownership.

2. Home-based vehicle travel is affected by the built environment to a lesser degree than by social or economic factors, as both home-based VHT and VMT models include only a single built environment variable of moderate significance. This may be attributable to the residential self-selection phenomenon.

3. Minimizing Mallow's C_p as a regression objective appears to create better models than maximizing the adjusted R^2, as fewer variables can be used with only negligible loss in explanatory power.

4. Principal component analysis can be used to examine the data structure of the explanatory variable space. For this study, it revealed that three principal components account for most (64%) of the variability in the explanatory variable space: the first is related to aspects of the transportation infrastructure, the second is related to aspects of population density and distribution, and the third is related to household demographics.

5. Power transformations to normality, when applied to explanatory variables, do not improve the modeling of mode choice, as the dependent mode choice variables are skewed asymmetrically. However, transformations such as the Yeo-Johnson family of power transformations do improve the modeling of more normally distributed variables, such as household VHT and VMT.

6. The models can easily be applied in a post-processor fashion to land use planning scenarios. For the hypothetical high-density residential development overviewed in

section 3.5, it was estimated that the proportion of a zone's trips taken by vehicle may be lowered by as much as 8% by re-developing low-density suburban housing to a higher density.

Many principles of smart growth, such as promoting mixed land use development, appear to be a valid way to encourage non-vehicle travel in the study area, as indicated by the statistical significance of diversity measures in the non-motorized and transit models. Moreover, high density development appears to encourage non-motorized travel, and dense street networks appear to encourage transit usage. As few built environment factors were present in the household home-based VHT and VMT models, the effects of smart growth planning on these measures of travel behavior is less visible.

5.2 Enhanced Four-step Travel Demand Forecasting Method

The most important finding of this part of the research pertains to the enhanced methodology developed. The study demonstrates that the traditional four-step method could be enhanced and made more sensitive to smart growth strategies. The enhanced travel demand forecasting method has many advantages over traditional method, particularly with respect to destination choice and mode choice, two of the most important components of travel demand forecasting process. The disaggregate destination choice and mode choice model used in this research allow incorporation of a multitude of socio-economic variables, and also zonal land use variables of both the two trip ends. The scenario studied show that it is applicable to extend the enhanced travel demand forecasting method beyond academic area and into practice.

This research also supports the claims that compact, mixed-use, pedestrian-friendly and transit-friendly designs can reduce vehicle trips, encourage non-motorized modes, decrease average trip length, and reduce daily VMT. Despite of having subjective factors in determining which variable to include in the model, the enduring statistical significance of some variables lend strong proof to the claim that smart growth strategies could reduce VMT.

Specifically, higher dissimilarity land uses can shorten the HBW and NHBW trips by increasing the probability of choosing destination TAZs less than 4.65 miles. More evenly distributed land use pattern (higher land use entropy) makes the HBW and HBShop trips shorter. Provision of transit infrastructure also has impacts on the destination choices of HBShop and HBSRO trips. Higher transit coverage rate is associated with shorter HBShop and HBSRO trips.

With respect to mode choice, compared with inter-zonal trips, intra-zonal trips' traffic mode choice is more sensitive to built-environmental changes, maybe because non-motorized modes are more likely to be available for short-distance intra-zonal trips than for longer inter-zonal trips. Higher population density and employment density both encourage non-motorized mode, while discourage auto travels in intrazonal trips. Provision of more convenient transit service increases the mode share of bus and rail for both intrazonal and interzonal trips.

Overall, the effects of the built environment factors on travel demand in the Greater Buffalo-Niagara area appear to be moderate at best. The findings are best summarized by the elasticities of VMT with respect to the land use variables quantifying built environment, resulting from applying the enhanced methodology, which revealed values falling in the range of 0.01 to 0.1 (see Table 5.1).

Table 5.1 Elasticities between VMT of Different Trip Purpose and Built Environment

	Dissimilarity index	Land use entropy	Transit coverage rate
VMT of HBW trips	-0.03	-0.075	
VMT of HBShop trips		-0.081	-0.1
VMT of HBSRO trips			-0.011
VMT of NHBW trips	-0.041		
VMT of NHBO trips			
Overall VMT	-0.01	-0.027	-0.013

Findings from the study thus support the claims that compact, mixed-use, pedestrian-friendly and transit-friendly designs can reduce vehicle trips, encourage non-motorized modes, decrease average trip length, and reduce daily VMT. Moreover, the study has developed two useful methodologies which can be applied to increase the sensitivity of current modeling tools toward assessing the likely impacts of proposed smart growth strategies.

5.3 Post Processor Method versus the Enhanced Four-Step Demand Forecasting Method

Each of the two approaches investigated in this study to increase the sensitivity of transportation planning models to the likely impacts of smart growth, have their own set of strengths and limitations. The post-processor method is quite straightforward, easy to implement, and very computationally efficient. However, the models developed would be specific to the geographic area from which the data used in developing the model were obtained. On the other hand, the enhanced four-step process defines a general framework which may be applied to any region. While the specific models would still need to be recalibrated for each region using that particular region household travel survey data, the enhanced framework is general enough to allow it to be applied to any region. Finally, while the data and computational requirements of the enhanced four-step process exceed those for the post-processor method, the results from the enhanced process are expected to be of higher accuracy.

6. REFERENCE

Ben-Akiva, M. and S. Lerman (1985). Discrete Choice Modeling: Theory and Application to Travel Demand. Cambridge, MIT Press.

Bento, A. M., M. L. Cropper, et al. (2005). "The Effects of Urban Spatial Structure on Travel Demand in the United States." Review of Economics and Statistics **87**(3): 466-478.

Bhat, C., A. Govindarajan, et al. (1998). "Disaggragate Attraction-End Choice Modeling Formulation and Empirical Analysis." Transportation research record **1645**: 60-68.

Boarnet, M. G. and M. G. Greenwald (2000). "Land Use, Urban Design, and Non-Work Travel: Reproducing for Portland, Oregon, Empirical Tests from Other Urban Areas " Transportation research record **1722**: 27-37.

Bradley, M. and P. Vovsha (2005). "A Model for Joint Choice of Daily Activity Pattern Types of Household Members." Transportation **32**: 545-571.

Cambridge_Systematics_Inc (1998). New Hampshire statewide travel demand model system- modal documentation report, prepared for the New Hampshire DOT, July.

Cao, X., P. L. Mokhtarian, et al. (2006). "Neighborhood Design and Vehicle Type Choice: Evidence from Northern California." transportation Research Part D **11**: 133-145.

Cao, X., P. L. Mokhtarian, et al. (2009). "The Relationship between the Built Environment and Nonwork Travel: A Case Study of Northern California." Transportation Research Part A **49**: 548-559.

Cervero, R. (1996). "Mixed Land-Uses and Communiting: Evidence From the American Housing Survey." Transportation Research Part A **30**(5): 361-377.

Cervero, R. (2002). "Built Environments and Mode Choice: toward a Normative Framework." transportation Research Part D **7**(4): 265-284.

Cervero, R. (2006). "Alternative Approaches to Modeling the Travel-Demand Impacts of Smart Growth." Journal of the American Planning Association **72**(3): 285-295.

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

Cervero, R. and C. Radisch (1996). "Travel Choices in Pedestrian versus Automobile Oriented Neighborhoods " Transport Policy **3**(3): 127-141.

Criteron_Planner&Engineers and Fehr&Peers_Associates (2001). INDEX 4D Method: A Quick-Response Method of Estimating Travel Impacts from Land-Use Changes Technical Memorandum, US Environmental Protection Agency.

DKS_Associates (2007). Assessment of Local Models and Tools For Analyzing Smart-growth Strategies, State of California Business, Transportation and Housing Agency, California Department of Transportation

Ewing, R. and R. Cervero (2001). "Travel and the Built Environment: A Synthesis." Transportation research record **1780**: 87-114.

Ewing, R., E. Dumbaugh, et al. (2001). "Internalizing Travel by Mixing Land Uses: Study of Master-Planning Communities in South Florida." Transportation research record **1780**: 115-120.

Frank, L. D. and G. Pivo (1994). "Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, and Walking." Transportation research record **1466**: 44-52.

Greenwald, M. J. (2006). "The Relationship between Land Use and Intrazonal Trip Making Behaviors: Evidence and Implications." transportation Research Part D **11**(6): 432-446.

Handy, S. L. (1996). "Urban Form and Pedestrian Choices: Study of Austin Neighborhoods." Transportation research record **1552**: 135-144.

Jonnalagadda, N., J. Freedman, et al. (2001). " Development of a Micro-simulation Activity-based Model for San Francisco: Destination and Mode Choice Models." Transportation research record **1777**: 25-30.

Kitamura, R., P. L. Mokhtarian, et al. (1997). "A Micro-analysis of Land Use and Travel in Five Neighborhoods in the San Francisco Bay Area." Transportation **24**: 125-158.

Kockelman, K. M. (1997). "Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from San Francisco Bay Area." Transportation research record **1607**: 116-125.

Loudon, W. R. and T. Parker (2008). "Modeling the Travel Impacts of Smart-Growth Strategies. ." The 87th Transportation Research Board Annual Meeting, CD-ROM.

McCann, J. (2003). Does New Urbanist/Neo Traditional Designs Deliver On Promises to Affect Modal Choice and Reduce Vehicle Miles Traveled Urban and Regional Planning. Blacksburg, Virginia Polytechnic Institute and State University. **Master:** 59.

Ortuzar, J. d. D. and L. G. Willumsen (2001). Modelling Transport. Chichester, John Wiley & Sons Ltd.

Potoglou, D. (2008). "Vehicle-type Choice and Neighbourhood Characteristics: An Empirical Study of Hamilton, Canada." transportation Research Part D **13**: 177-186.

Rajamani, J., C. R. Bhat, et al. (2003). "Assessing the Impact of Urban Form Measures on Nonwork Trip Mode Choice after Controlling for Demographic and Level-of-service Effects." Transportation research record **1831**: 158-165.

Shiftan, Y. (1998). " A Practical Approach to Model Trip Chaining." Transportation research record **1645**: 17-23.

Shiftan, Y. and M. Ben-Akiva (2010). "A Practical Policy-sensitive, Activity-based, Travel-demand Model." The Annals of Regional Science.

US_EPA. (Oct. 14 2010). "Smart Growth Principles." from http://www.epa.gov/smartgrowth/about_sg.htm.

Vega, A. and A. Reynolds-Frighan (2009). "A Methodological Framework for the Study of Residential Location and Travel-to-work Mode Choice under Central and Suburban Employment Destination Patterns." Transportation Research Part A **43**: 401-419.

Zhang, M. (2006). "Travel Choice with No Alternative: Can Land Use Reduce Automobile Dependence." Journal of Planning Education and Research **25**: 311-326.

7. APPENDICES

Appendix A

Appendix A

Travel behavior variables

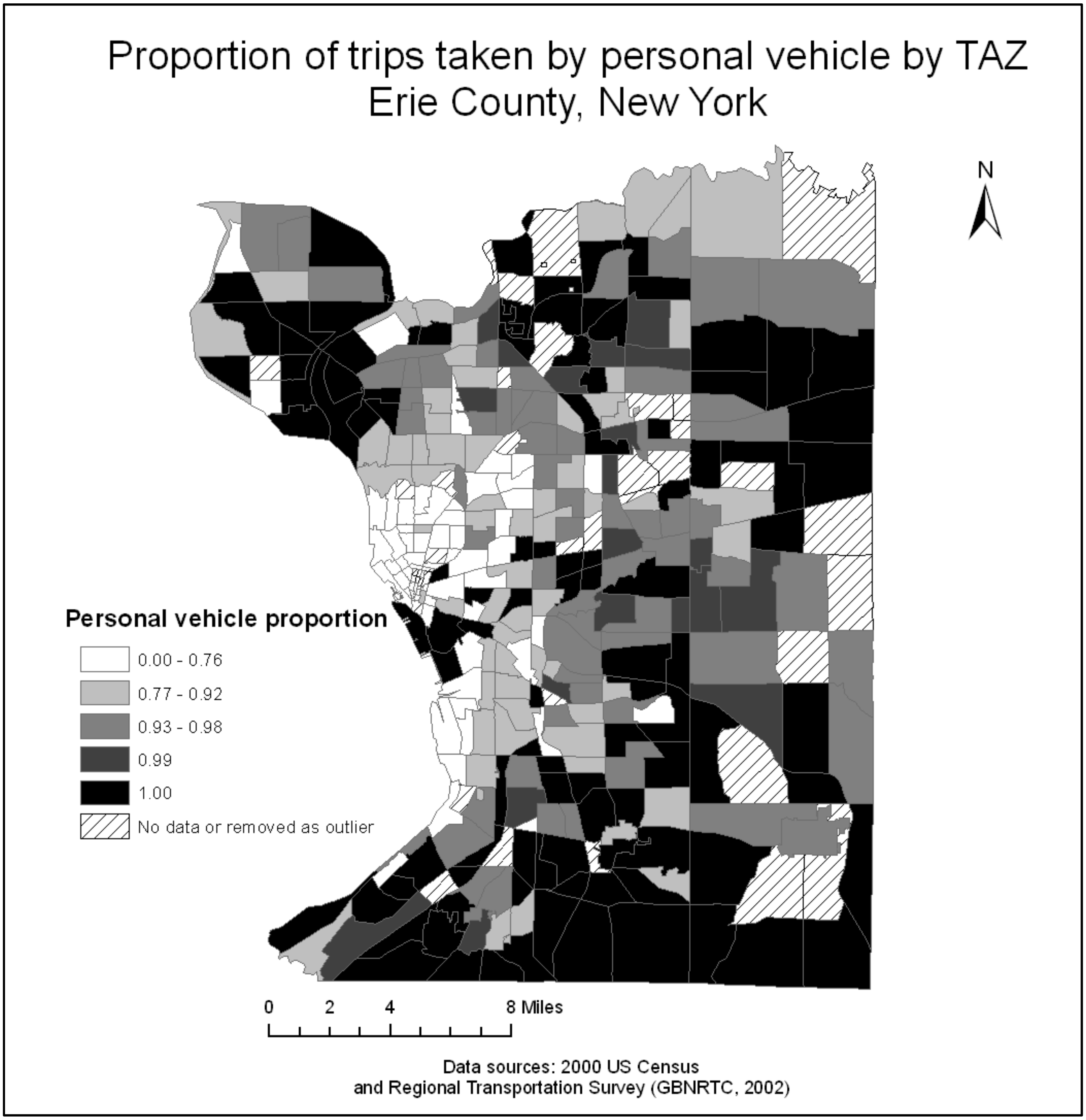


Figure A1: Chloropleth map of vehicle trip proportion by TAZ

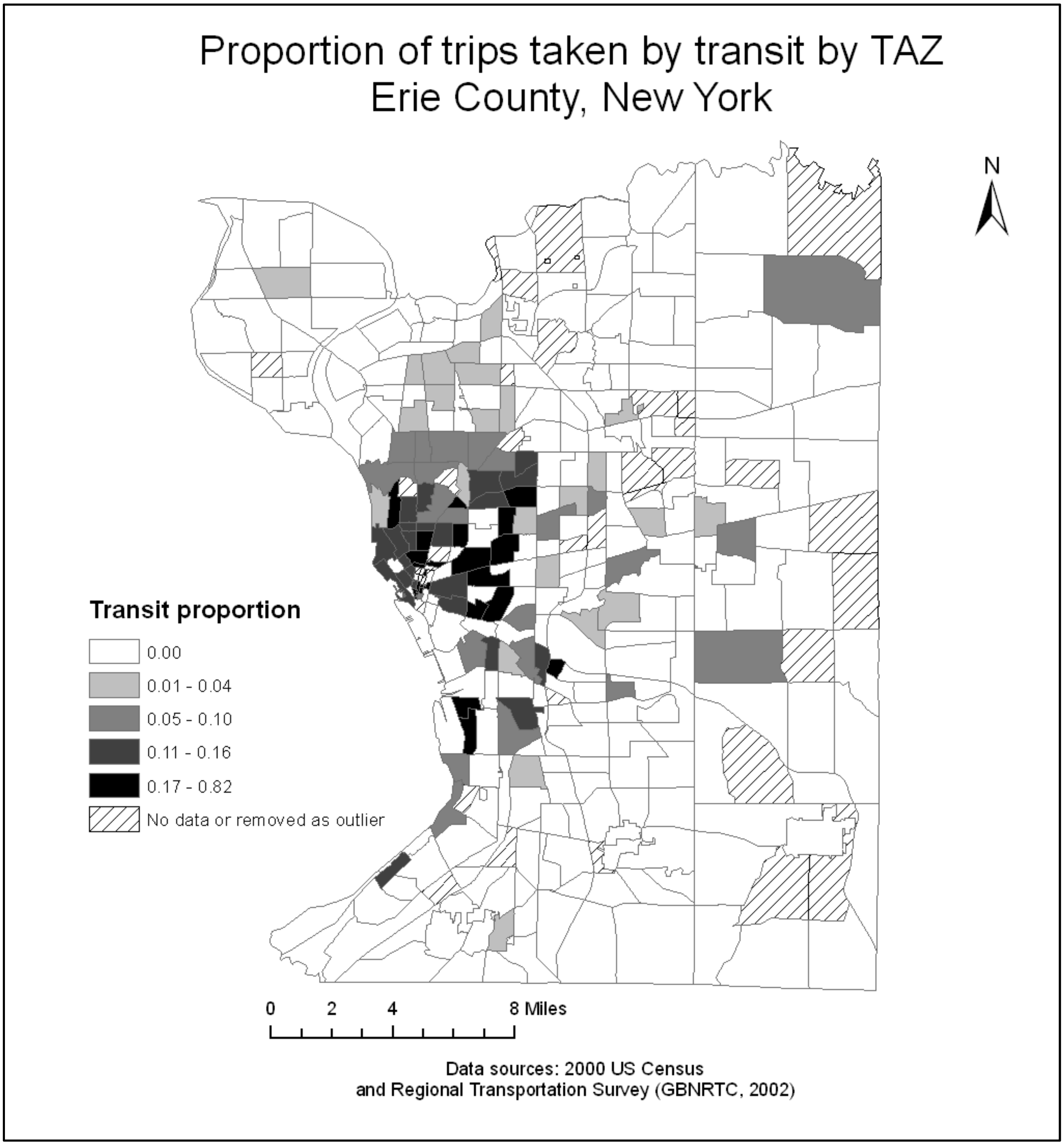


Figure A2: Chloropleth map of transit trip proportion by TAZ

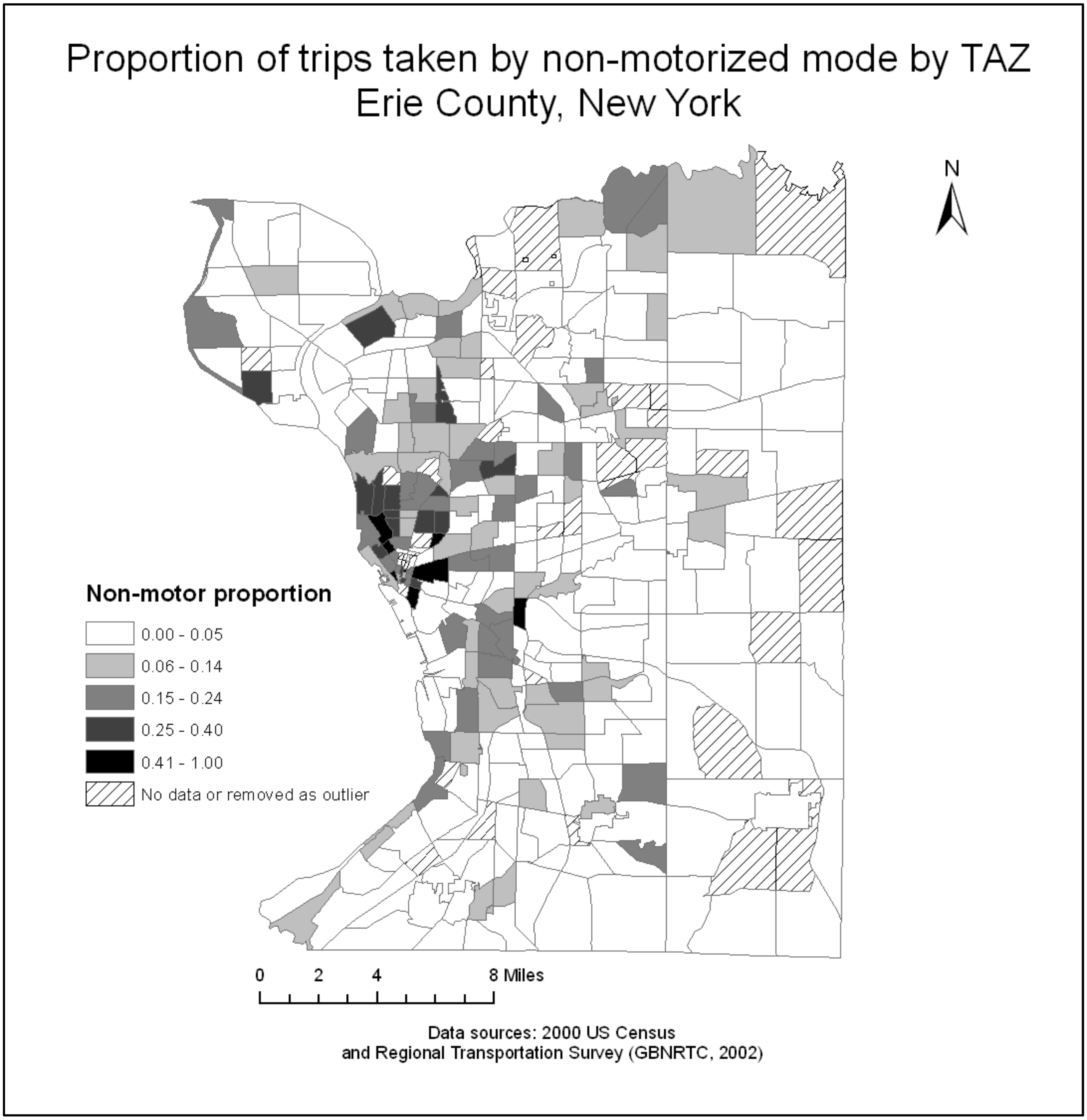


Figure A3: Chloropleth map of non-motorized trip proportion by TAZ

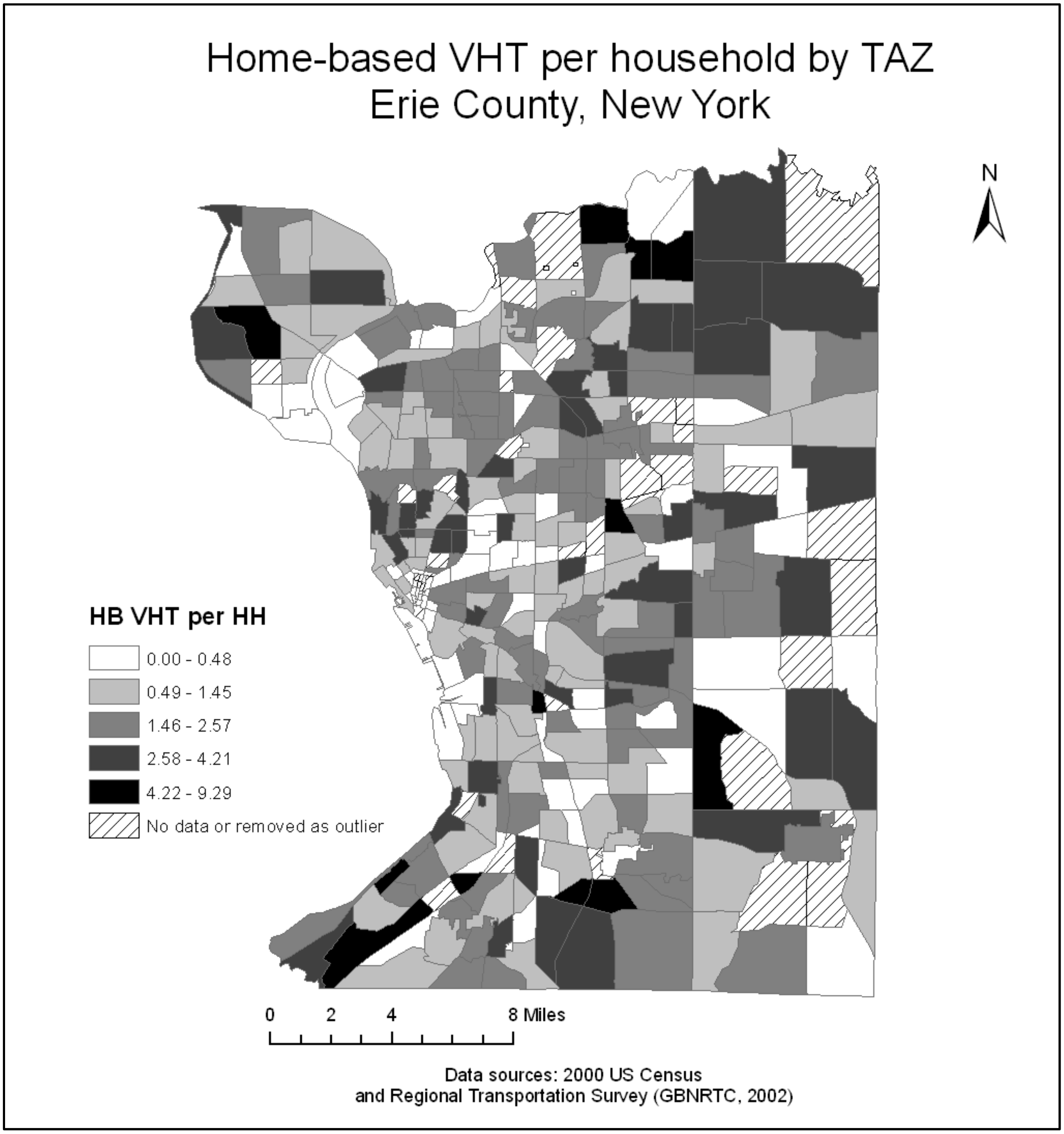


Figure A4: Chloropleth map of home-based daily VHT per household by TAZ

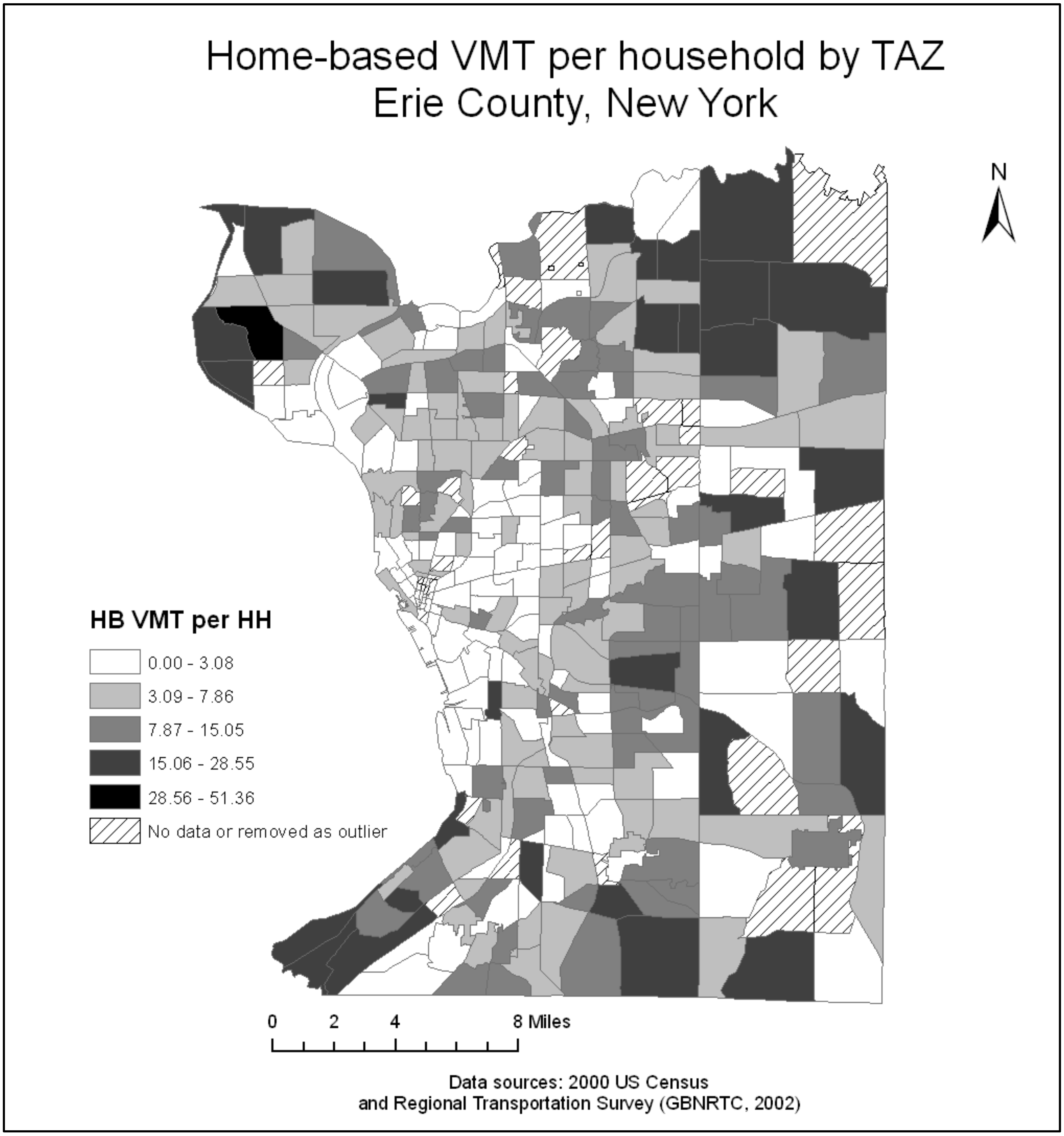


Figure A5: Chloropleth map of home-based VMT per household by TAZ

Table A1: Summary statistics for travel behavior variables (mean, median, and quartiles)

NonMSplit	TransitSplit	VehSplit	HB_VHT_pHH	HB_VMT_pHH
Min. :0.00000	Min. :0.00000	Min. :0.0000	Min. :0.0000	Min. : 0.000
1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.8226	1st Qu.:0.5799	1st Qu.: 1.861
Median :0.02669	Median :0.00000	Median :0.9577	Median :1.3705	Median : 4.900
Mean :0.07724	Mean :0.03817	Mean :0.8655	Mean :1.5632	Mean : 6.413
3rd Qu.:0.10927	3rd Qu.:0.02102	3rd Qu.:1.0000	3rd Qu.:2.3140	3rd Qu.: 9.292
Max. :1.00000	Max. :0.81620	Max. :1.0000	Max. :9.2872	Max. :51.358

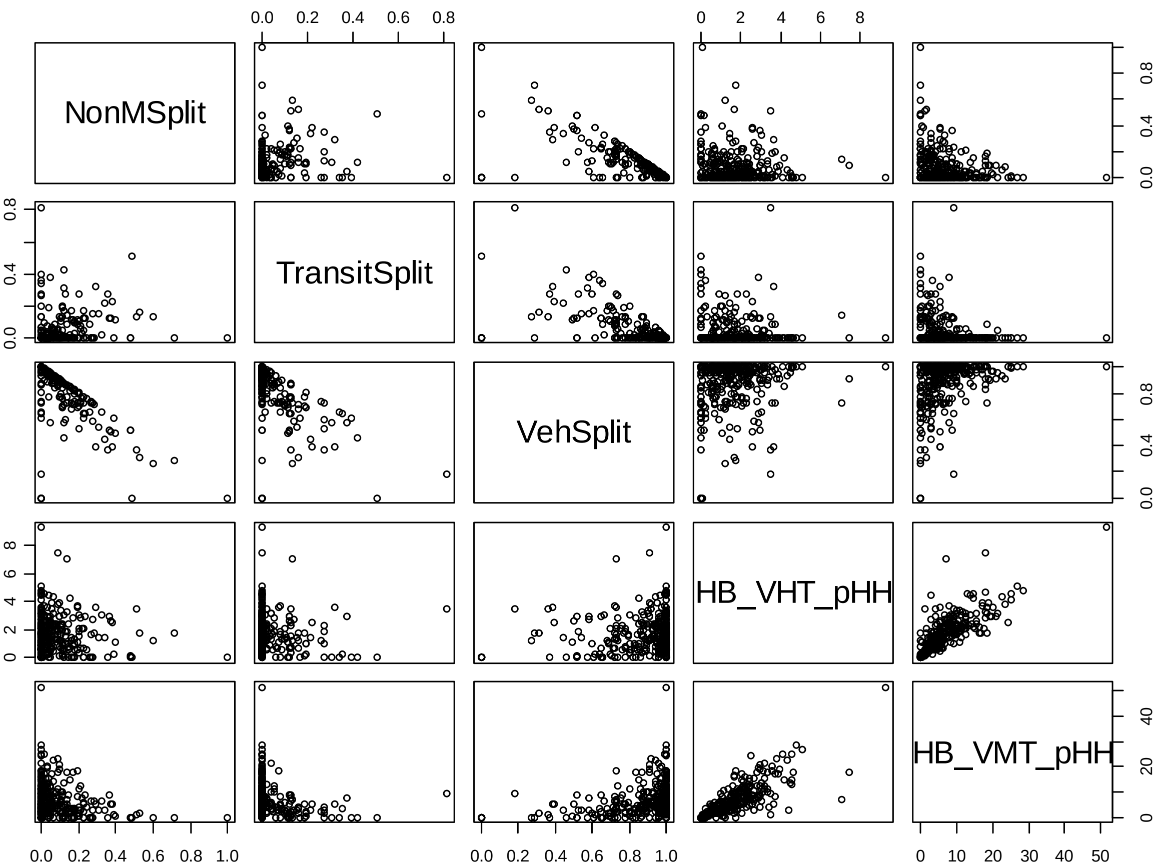


Figure A6: Matrix plot of travel behavior variables

As all modes included in the travel survey were categorized as non-motorized, transit, or vehicle, the sum of all three mode proportions for each zone is one. This constraint is clearly visible as the diagonal distribution of the mode choice scatterplots (upper left). Negative correlations between non-motorized or transit mode choice and the measures of vehicle travel are also clearly visible. Most zones appear to rely entirely on vehicle travel, as confirmed by the quartiles in Table A1, above, and in Figure A7, below.

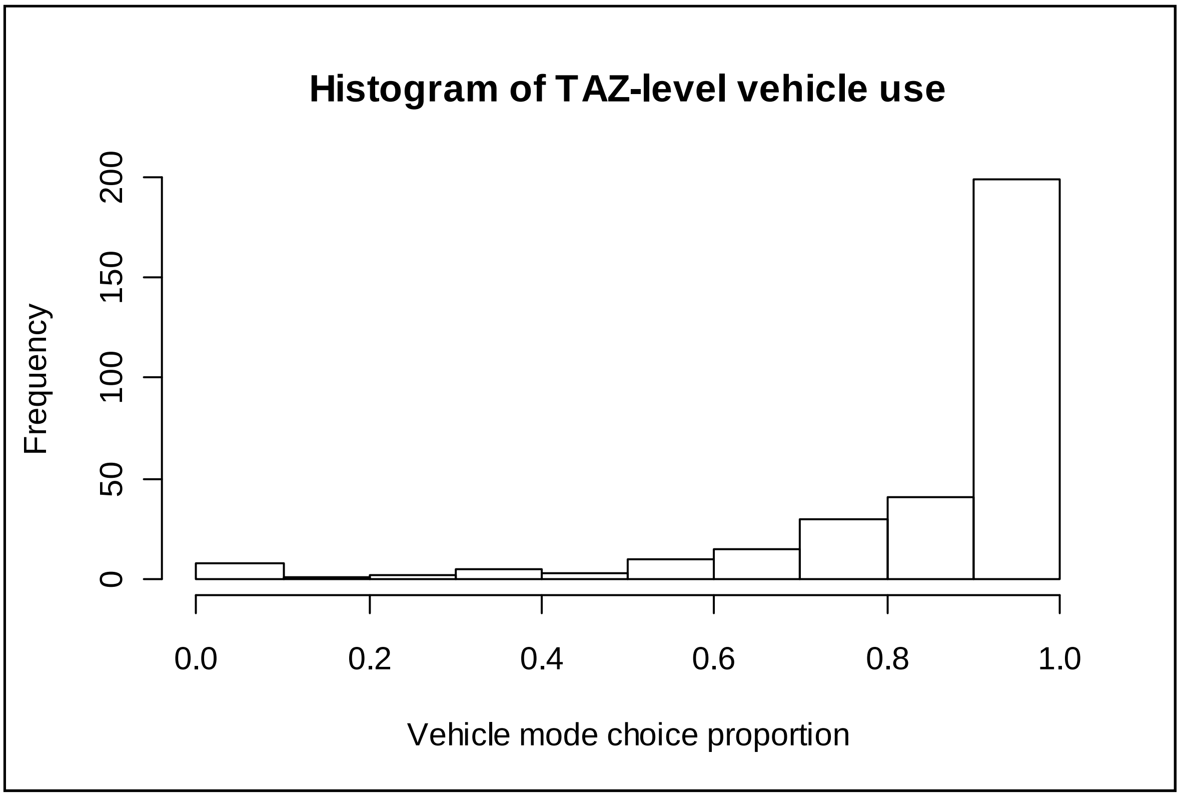


Figure A7: Personal vehicle dependence in Erie County

As seen above, roughly two-thirds of the study area TAZs are heavily reliant on personal vehicle travel, which accounts for 90% or more of the trips taken in these zones. Spatial variation in vehicle dependence can be seen in Figure A1. Distributions of the other two mode choices, non-motorized and transit, can be found in Figure A8 and Figure A9, respectively.

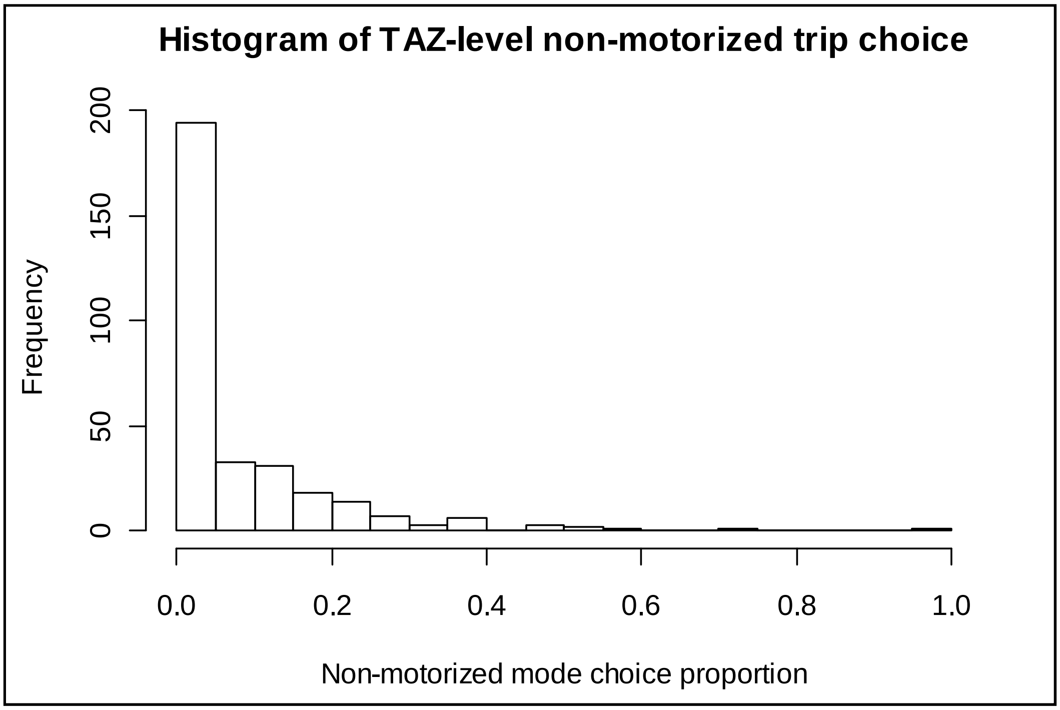


Figure A8: Distribution of non-motorized trip proportions

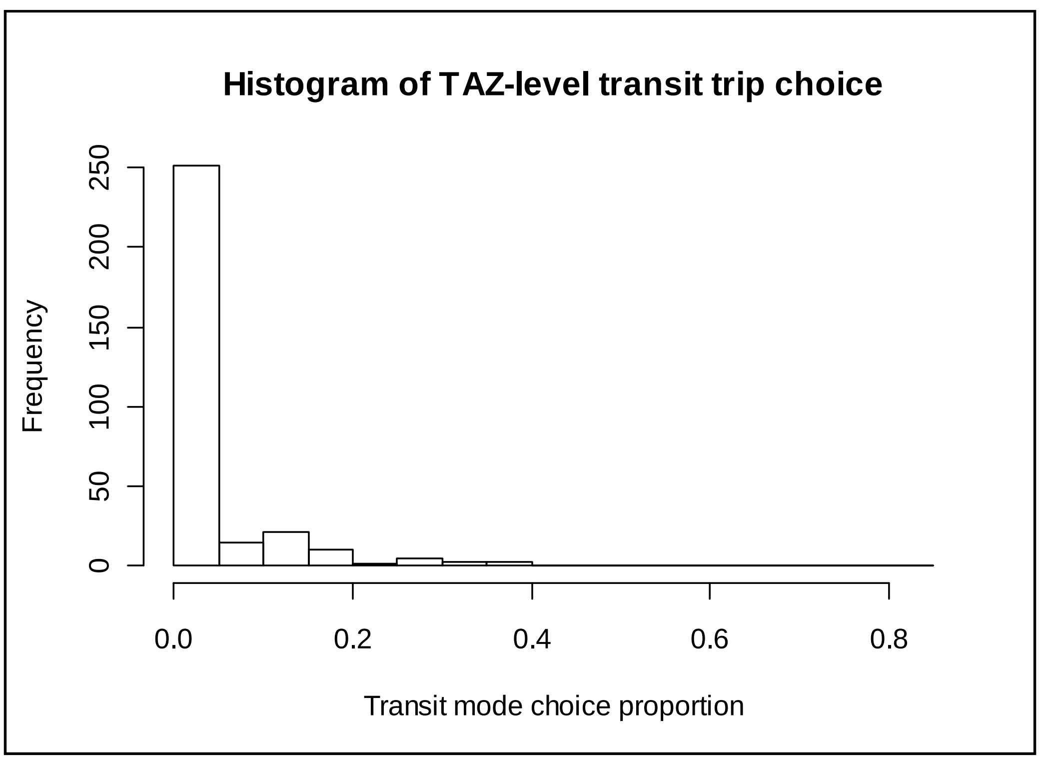


Figure A9: Distribution of transit trip proportions

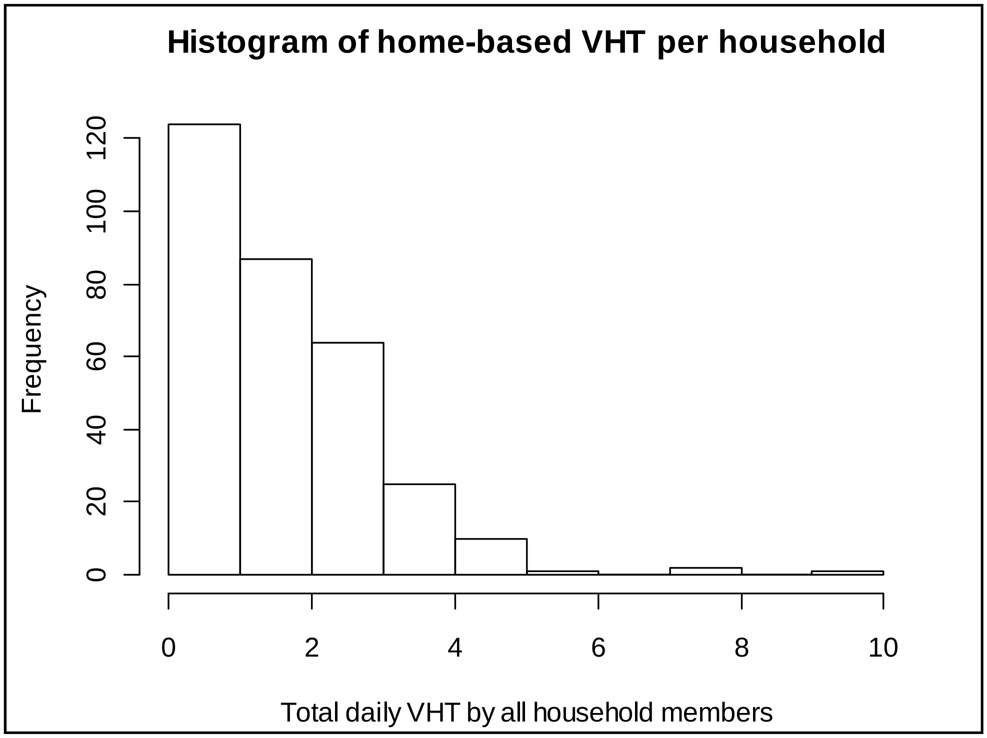


Figure A10: Distribution of home-based VHT per household

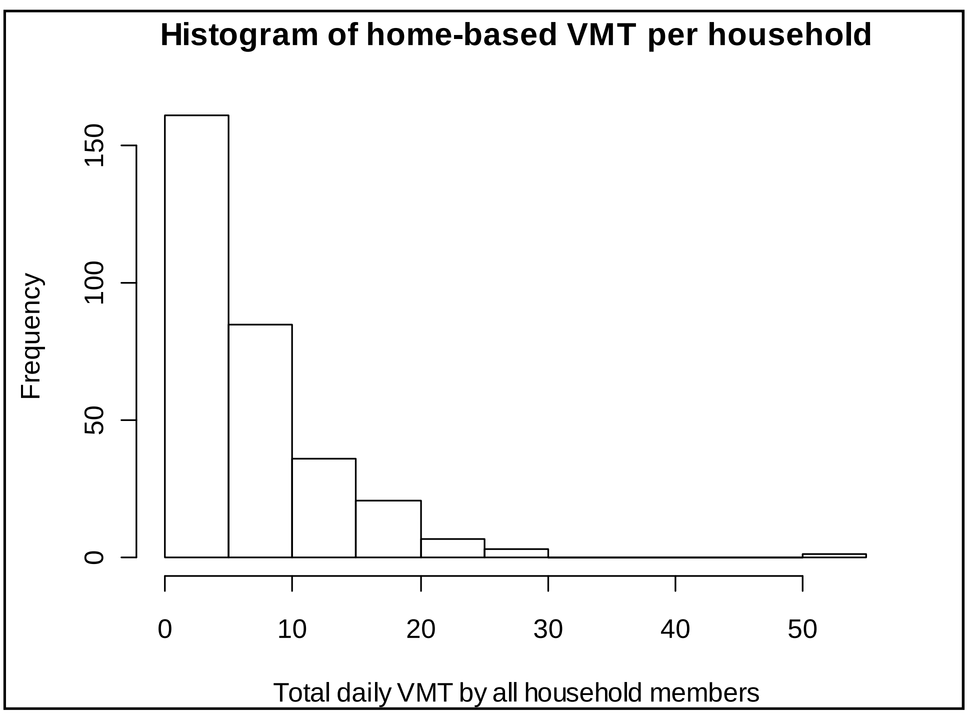


Figure A11: Distribution of home-based VMT per household

Appendix B

Appendix B
Explanatory variables

Prior to modeling, a much larger set of variables (including transformed variables) was tested for correlation with the dependent variables. Many with low correlations were discarded, and others were discarded for quantifying the same effect as another variable (for example, several ways of computing population density were tested, and only the best was kept). Table XXXXX from Chapter 3: Methodology shows the variables eventually included in linear models. Table B1, below, shows the variable space initially used in linear modeling, including variables not present in the final models, along with expanded descriptions of each variable.

Table B1: Explanatory variable descriptions

Demographics	MHHI_2000	Median household income from the 2000 census
	HHVEH	Number of vehicles owned by the household
	HWORK	Number of employed persons in the household
	HSTUD	Number of students in the household
	HHSIZ	Total residents of the household. Household vehicles, workers, students and size were self-reported by survey participants.
Built environment variables	PopDensity	Population density, computed as population divided by the total area of the zone. Other measures of population density were attempted, such as population over developed area and population over residential area; these measures were discarded prior to modeling due to having slightly worse correlations with the dependent variables.
	EmpDensity	Employment density, computed as persons employed in the zone divided by the total area of the zone.
	ResComm Balance	These balance-based variables were computed using the balance formula reviewed in Methodology. This variable is the balance of residential and commercial land use area. Many

		other land use balance measures were attempted and discarded prior to modeling.
	SingleResOtherRes Balance	The balance of single-family residential area to all other types of residential area (including double-family, triple-family, apartments, and rural housing). The variable is intended to quantify mix of housing types, thought to shorten mean commute times by allowing people of all housing preferences to live near their workplace.
	ResCommun Balance	The balance of residential and community land use area. As community land uses include schools, health facilities, churches, and other considerable trip attractors, it is thought that mixing community services into residential areas may reduce VMT.
	CommunComm Balance	The balance of community and commercial land use area. Mixing these may encourage trip chaining.
	ApartmentOtherRes Balance	The balance of apartments to all other types of residential land uses. As above, diverse available housing options may reduce commute times. As few TAZs have significant area devoted to apartments, significant variation only exists among urban TAZs.
	SNDbyParcelArea	Street network density, computed as the total length of street divided by the parcel-covered area of the zone. Other methods for computing street network density were attempted, including dividing by the total area of the zone and by the developed area of the zone; these variables were discarded in favor of per-parcel- area.
	JunctionD byStreetLength	Junction point density, computed as the number of junctions (modeled as nodes in ArcGIS Network Analyst, including intersections and interchanges) divided by the total length of street. Other methods of computing junction density included

		junctions per total area and junctions per developed area; these were discarded. This variable describes the same aspect of the built environment as junction kernel density. As the variables were not too highly correlated with one another (*p*=0.74), both were included in the variable set for linear modeling, and at most one of these two variables was allowed in each model.
	StreetProx400ft	Street network proximity, computed as the proportion of parcel area within 400ft of any street. Similar to street network density, which was used in favor of this variable in the final linear models. Other distances were attempted; 400ft was found to maximize variability (giving dense urban zones a value close to 1, sparse rural zones a value close to 0, and most zones well-distributed in between) and correlation with the dependent variables.
	TransitPoint	Transit point density, computed as the total number of transit stops (for both buses and rail) divided by total area. This variable was eventually discarded in favor of transit kernel density.
	TransitKernel	Transit kernel density, computed as the mean transit kernel value for the zone. All transit points were given the same weight and kernel radius, set at 100m. The concept of kernel density is explained in the Methodology section.
	JunctionKernel	Junction kernel density, computed as the mean junction kernel value for the zone. Each junction was given the same weight and kernel radius, set at 100m. This variable was expected to partially account for the pedestrian friendliness of a zone, as short blocks are expected to encourage pedestrian travel.
	RoadKernel	Road kernel density, computed as the mean road kernel value for the zone. All roads were given the same weight and kernel radius, set at 100m. This is the only line-kernel variable – all

		other kernel densities are point-kernels.
	EmpKernelWeighted	Employment kernel density, computed as the mean employment kernel value for the zone. Each place of employment was given a kernel function weighted for the number of employees and a radius of 0.5 miles. This variable was transformed with the *log* function prior to inclusion in the regression variable set so as to improve its distribution. Prior to applying the logarithm, the variable spanned many orders of magnitude. Its correlation with the dependent variables was significantly improved.
	MeanFZ1PD	The number of fare zone 1 transit stops per unit area. At the time the travel survey was conducted, the Buffalo area metro system used multiple fare zones, with zone 1 as the most urban and zone 4 as the most rural. It was found that TAZs that were at least partially contained in zone 1 had significantly higher transit usage than those in zones 2, 3 or 4, due to extra fees incurred when travelling between zones and better transit infrastructure in zone 1.
	Dissim100m	Mean dissimilarity index for all cells in the zone. More on this variable can be found in methodology. Another dissimilarity index, based on a smaller, 50m grid was attempted. As many parcels are smaller than 100m, it was thought that 50m would better capture land use mixing. However, the 100m-based dissimilarity index was found to be better correlated with travel behavior, and the 50m-based index was discarded.
	ResPercent	Proportion of parcel area designated as residential. Low correlation with population density (p=0.26).
	CommercePercent	Proportion of parcel area designated as commercial. Correlated with employment density (p=0.49).
	EmpPercent	Proportion of parcel area designated as employment

		(commercial, community or industrial, excluding apartments & community parcels that are relatively undeveloped).
	DevPercent	Proportion of parcel area that is developed – that is, not vacant, unknown, agricultural, forest, or relatively undeveloped community parcels.
	HighDIndex1	Crafted as an indicator of urban areas, this index is the percentage of parcel area classified as apartments, two- or three-family houses, offices, retail, or multi-use. These land uses were observed to be negatively correlated to vehicle use (both vehicle mode choice and VMT). Well correlated with population density (p=0.70) but not employment density (p=0.22).
	ResPercentDev	Proportion of developed parcel area classified as residential.
	EmpPercentDev	Proportion of developed parcel area classified as employment. Along with ResPercentDev, not in any final linear models.

Table B2: Summary statistics for built environment variables (mean, median, and quartiles)

MHHI_2000	HHVEH	HWORK	HSTUD	HHSIZ
Min. : 0 1st Qu.: 28919 Median : 41368 Mean : 42375 3rd Qu.: 53986 Max. :113755	Min. :0.000 1st Qu.:0.958 Median :1.447 Mean :1.384 3rd Qu.:2.000 Max. :4.000	Min. :0.000 1st Qu.:0.540 Median :1.004 Mean :1.013 3rd Qu.:1.439 Max. :4.000	Min. :0.0000 1st Qu.:0.0000 Median :0.4895 Mean :0.6118 3rd Qu.:0.9526 Max. :4.0559	Min. :0.000 1st Qu.:1.629 Median :2.187 Mean :2.100 3rd Qu.:2.738 Max. :7.000
PopDensity	**EmpDensity**	**ResPercent**	**CommercePercent**	**EmpPercent**
Min. : 0 1st Qu.: 1579 Median : 4705 Mean : 6602 3rd Qu.: 8955 Max. :50925	Min. : 18.24 1st Qu.: 754.39 Median : 2217 Mean : 8731 3rd Qu.: 5038 Max. :340339	Min. :0.0000 1st Qu.:0.2694 Median :0.4488 Mean :0.4432 3rd Qu.:0.6267 Max. :0.9164	Min. :0.00000 1st Qu.:0.03493 Median :0.09192 Mean :0.12751 3rd Qu.:0.16580 Max. :0.72266	Min. :0.00000 1st Qu.:0.07716 Median :0.15593 Mean :0.20223 3rd Qu.:0.25765 Max. :0.97232
HighDIndex1	**ResComm Balance**	**SingleRes OtherResBalance**	**ResCommun Balance**	**CommunComm Balance**
Min. :0.00000 1st Qu.:0.03423 Median :0.08534 Mean :0.12914 3rd Qu.:0.19066 Max. :0.62379	Min. :0.0000 1st Qu.:0.2742 Median :0.5786 Mean :0.5350 3rd Qu.:0.7896 Max. :0.9995	Min. :0.0000 1st Qu.:0.3766 Median :0.4949 Mean :0.4481 3rd Qu.:0.5577 Max. :0.6099	Min. :0.0000 1st Qu.:0.2514 Median :0.5725 Mean :0.5225 3rd Qu.:0.8141 Max. :0.9942	Min. :0.0000 1st Qu.:0.1601 Median :0.4717 Mean :0.4643 3rd Qu.:0.7421 Max. :0.9841
Apartment OtherResBalance	**Dissim100m**	**SNDbyParcelArea**	**TransitKernel**	**JunctionKernel**
Min. :0.00000 1st Qu.:0.00000 Median :0.06417 Mean :0.16498 3rd Qu.:0.21439 Max. :0.98223	Min. :0.01756 1st Qu.:0.22552 Median :0.27504 Mean :0.30380 3rd Qu.:0.35668 Max. :0.80555	Min. : 1.965 1st Qu.: 11.095 Median : 19.668 Mean : 25.523 3rd Qu.: 32.164 Max. :190.472	Min. : 0.000 1st Qu.: 2.236 Median : 8.390 Mean : 110.602 3rd Qu.: 30.490 Max. :3704.965	Min. : 2.098 1st Qu.: 31.939 Median : 77.678 Mean : 87.072 3rd Qu.:123.928 Max. :589.747
RoadKernel	**MeanFZ1PD**			
Min. : 0.8849 1st Qu.: 5.9221 Median :10.8498 Mean :12.1290 3rd Qu.:17.2323 Max. :42.3906	Min. : 0.000 1st Qu.: 0.000 Median : 0.000 Mean : 3.113 3rd Qu.: 2.429 Max. :62.621			

Note that, in Table B2, mean values are not weighted for TAZ size. Thus, mean employment density is higher than mean population density, despite the study area having more people than employed people. This is an example of Simpson's paradox.

Table B3: Pearson correlation matrix, dependent and independent variables

	NonMSplit	TransitSplit	VehSplit	HB_VHT_pHH	HB_VMT_pHH
MHHI_2000	-0.32	-0.43	0.49	0.33	0.49
HHVEH	-0.3	-0.36	0.49	0.45	0.54
HWORK	-0.09	-0.17	0.26	0.44	0.47
HSTUD	0.21	-0.02	-0.03	0.48	0.39
HHSIZ	0.03	-0.11	0.19	0.58	0.52
PopDensity	0.5	0.38	-0.39	-0.02	-0.23
EmpDensity	0.21	0.38	-0.27	-0.19	-0.18
ResCommBalance	0.07	-0.09	0.11	0.23	0.09
SingleResOtherResBalance	-0.35	-0.48	0.48	0.24	0.31
ResCommunBalance	0.01	-0.06	0.16	0.14	0.01
CommunCommBalance	0.11	0.18	-0.04	-0.01	-0.03
ApartmentOtherResBalance	0.12	0.2	-0.2	-0.1	-0.16
SNDbyParcelArea	0.44	0.56	-0.45	-0.18	-0.28
JunctionDbyStreetLength	0.28	0.33	-0.27	-0.1	-0.28
StreetProx400ft	0.34	0.3	-0.21	0	-0.2
TransitPoint	0.32	0.44	-0.36	-0.24	-0.22
TransitKernel	0.29	0.45	-0.35	-0.22	-0.2
JunctionKernel	0.46	0.53	-0.42	-0.13	-0.3
RoadKernel	0.48	0.51	-0.41	-0.09	-0.29
EmpKernelWeighted	0.32	0.48	-0.37	-0.23	-0.24
MeanFZ1PD	0.38	0.55	-0.43	-0.23	-0.25
Dissim100m	0.16	0.46	-0.22	-0.19	-0.27
ResPercent	0.01	-0.17	0.2	0.3	0.25
CommercePercent	0.23	0.37	-0.24	-0.28	-0.34
EmpPercent	0.13	0.36	-0.34	-0.3	-0.34
DevPercent	0.16	0.15	-0.1	0.06	-0.04
HighDIndex1	0.42	0.41	-0.35	-0.08	-0.26
ResPercentDev	-0.05	-0.26	0.26	0.32	0.32
EmpPercentDev	0.06	0.3	-0.28	-0.32	-0.37

In Table B3, above, many of the household demographic variables (the first five rows) are well-correlated to the measures of travel behavior. All are positively correlated with vehicle use. Population density, transit kernel density, and junction kernel density – all of which are related to density – appear to be good candidate predictors for mode choice. The balance of single-family housing to all other types of residences (denoted by the variable SingleResOtherResBalance) appears to be the best diversity-related variable, and is positively correlated with vehicle use. This is likely due to the prevalence of double-family housing in suburban areas. Street network density, which is related to both density and design, is a surprisingly good indicator of both non-motorized travel and transit usage. The dissimilarity index was developed by Cervero and Kockelman (1997) to be an indicator of pedestrian travel, as it was thought that mixed land uses encourage non-vehicle work trips. However, for this study, dissimilarity of land uses was found to be only a weak indicator of non-motorized travel. A larger version of this matrix with more explanatory variables was used to screen for poorly-correlated variables to discard, and to select the best method of computing variables such as population density, for which there were several ways to define area (such as total area, parcel-covered area, and area classified as residential).

Table B4: Spearman correlation matrix, dependent and independent variables

	NonMSplit	TransitSplit	VehSplit	HB_VHT_pHH	HB_VMT_pHH
MHHI_2000	-0.32	-0.52	0.51	0.31	0.49
HHVEH	-0.23	-0.36	0.41	0.49	0.64
HWORK	0	-0.11	0.14	0.52	0.58
HSTUD	0.14	0.07	-0.04	0.62	0.51
HHSIZ	0.06	-0.08	0.08	0.69	0.65
PopDensity	0.47	0.51	-0.48	0.12	-0.06
EmpDensity	0.33	0.49	-0.43	-0.21	-0.38
ResCommBalance	0.17	0.08	-0.07	0.29	0.19
SingleResOtherResBalance	-0.18	-0.43	0.35	0.21	0.35
ResCommunBalance	0.18	0.09	-0.09	0.2	0.13
CommunCommBalance	0.24	0.25	-0.21	0.07	0.04
ApartmentOtherResBalance	0.2	0.3	-0.27	0.05	-0.04
SNDbyParcelArea	0.48	0.53	-0.52	-0.02	-0.21
TransitKernel	0.5	0.59	-0.56	-0.11	-0.32
JunctionKernel	0.51	0.56	-0.56	-0.01	-0.2
RoadKernel	0.52	0.54	-0.54	0.01	-0.18
MeanFZ1PD	0.49	0.61	-0.59	-0.17	-0.34
Dissim100m	0.07	0.29	-0.16	-0.16	-0.27
ResPercent	0.13	-0.01	0.02	0.36	0.35
CommercePercent	0.21	0.36	-0.26	-0.2	-0.34
EmpPercent	0.16	0.32	-0.28	-0.24	-0.36
HighDIndex1	0.43	0.53	-0.45	-0.01	-0.19

Explanatory variables not found in any of the finalized regression models were omitted from the above matrix.

Table B5: Pearson correlation matrix, dependent variables and Yeo-Johnson power transformed independent variables

	NonMSplit	TransitSplit	VehSplit	HB_VHT_pHH	HB_VMT_pHH
MHHI_2000	-0.32	-0.44	0.53	0.35	0.48
HHVEH	-0.29	-0.36	0.49	0.46	0.55
HWORK	-0.07	-0.17	0.27	0.48	0.5
HSTUD	0.15	-0.04	0.04	0.54	0.45
HHSIZ	0.02	-0.11	0.22	0.59	0.53
PopDensity	0.41	0.33	-0.26	0.04	-0.19
EmpDensity	0.31	0.46	-0.36	-0.28	-0.43
ResCommBalance	0.07	-0.09	0.11	0.23	0.1
SingleResOtherResBalance	-0.31	-0.46	0.43	0.21	0.29
ResCommunBalance	0.01	-0.06	0.16	0.15	0.01
CommunCommBalance	0.11	0.18	-0.04	-0.01	-0.04
ApartmentOtherResBalance	0.16	0.16	-0.18	-0.06	-0.16
SNDbyParcelArea	0.44	0.47	-0.36	-0.11	-0.31
TransitKernel	0.45	0.53	-0.43	-0.2	-0.38
JunctionKernel	0.43	0.46	-0.36	-0.1	-0.31
RoadKernel	0.43	0.44	-0.32	-0.06	-0.27
MeanFZ1PD	0.52	0.54	-0.49	-0.17	-0.31
Dissim100m	0.13	0.42	-0.17	-0.17	-0.25
ResPercent	0	-0.18	0.22	0.32	0.27
CommercePercent	0.2	0.33	-0.19	-0.25	-0.36
EmpPercent	0.12	0.32	-0.28	-0.25	-0.36
HighDIndex1	0.36	0.38	-0.28	-0.09	-0.28

Ideas for figures:
Study area land uses
Study area dissimilarity cells
Dissimilarity values (0-8)
Population, employment density
Road, transit, junction kernel density
Land use by type (in square feet and percentage for whole study area)

Appendix C

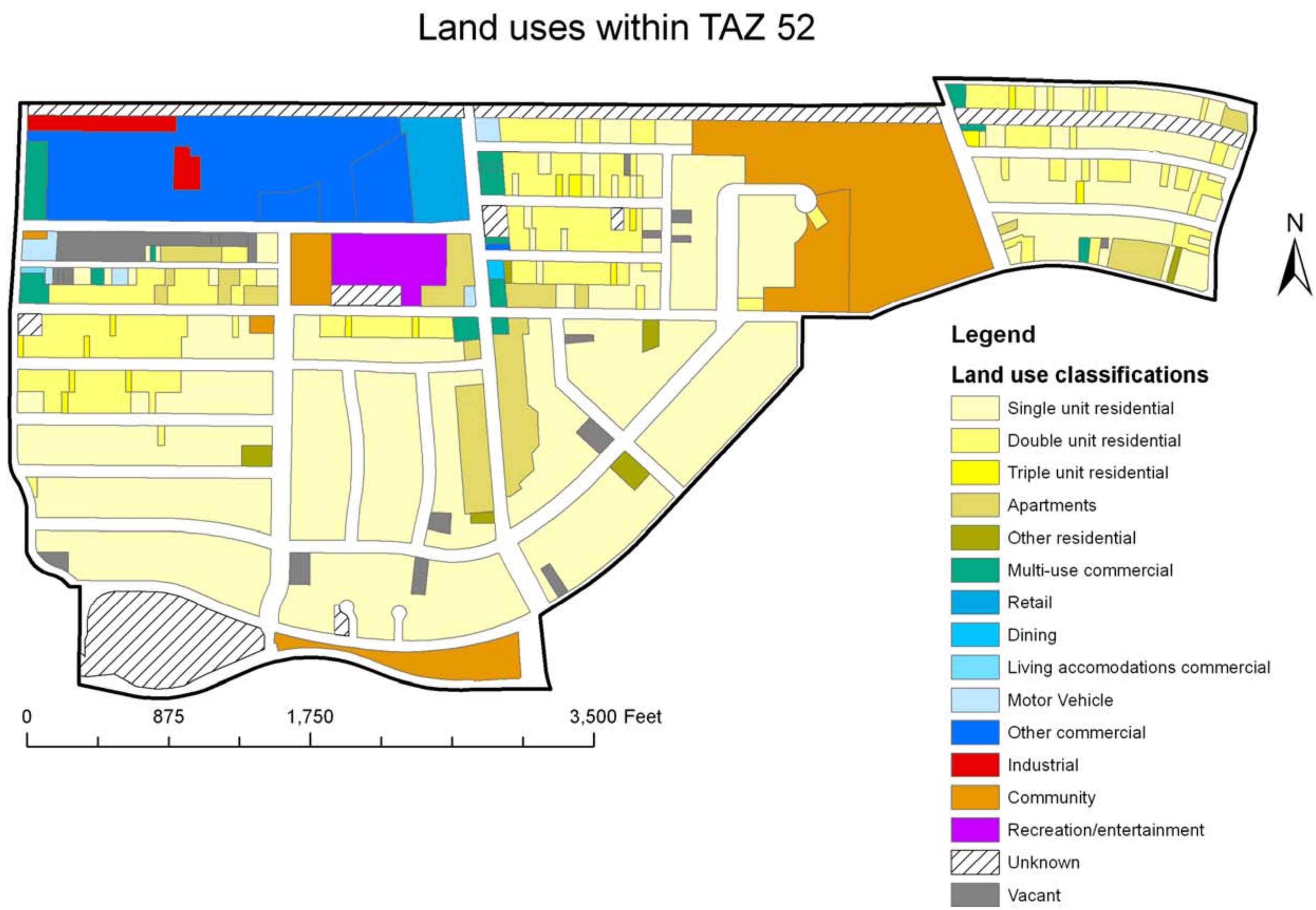


Figure C1: Land uses within TAZ 52

Parcels of the same type were aggregated; thus, individual parcel boundaries are not shown. Land use classifications are defined by the New York State Assessors' Manual (New York State Office of Real Property Services, 2006).

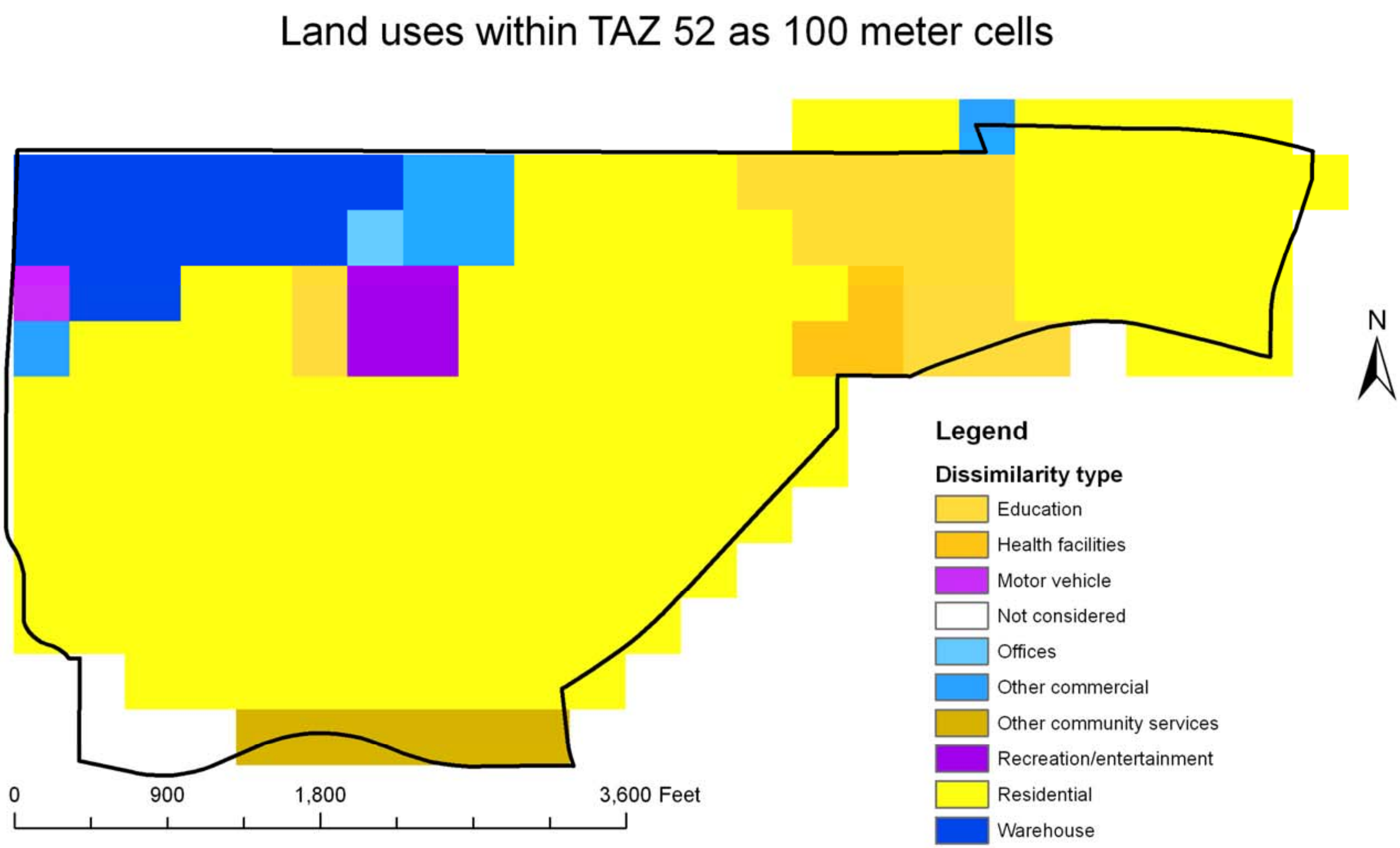


Figure C2: Land uses within TAZ 52 as 100 meter cells, as used for dissimilarity index computations

More bla goes here

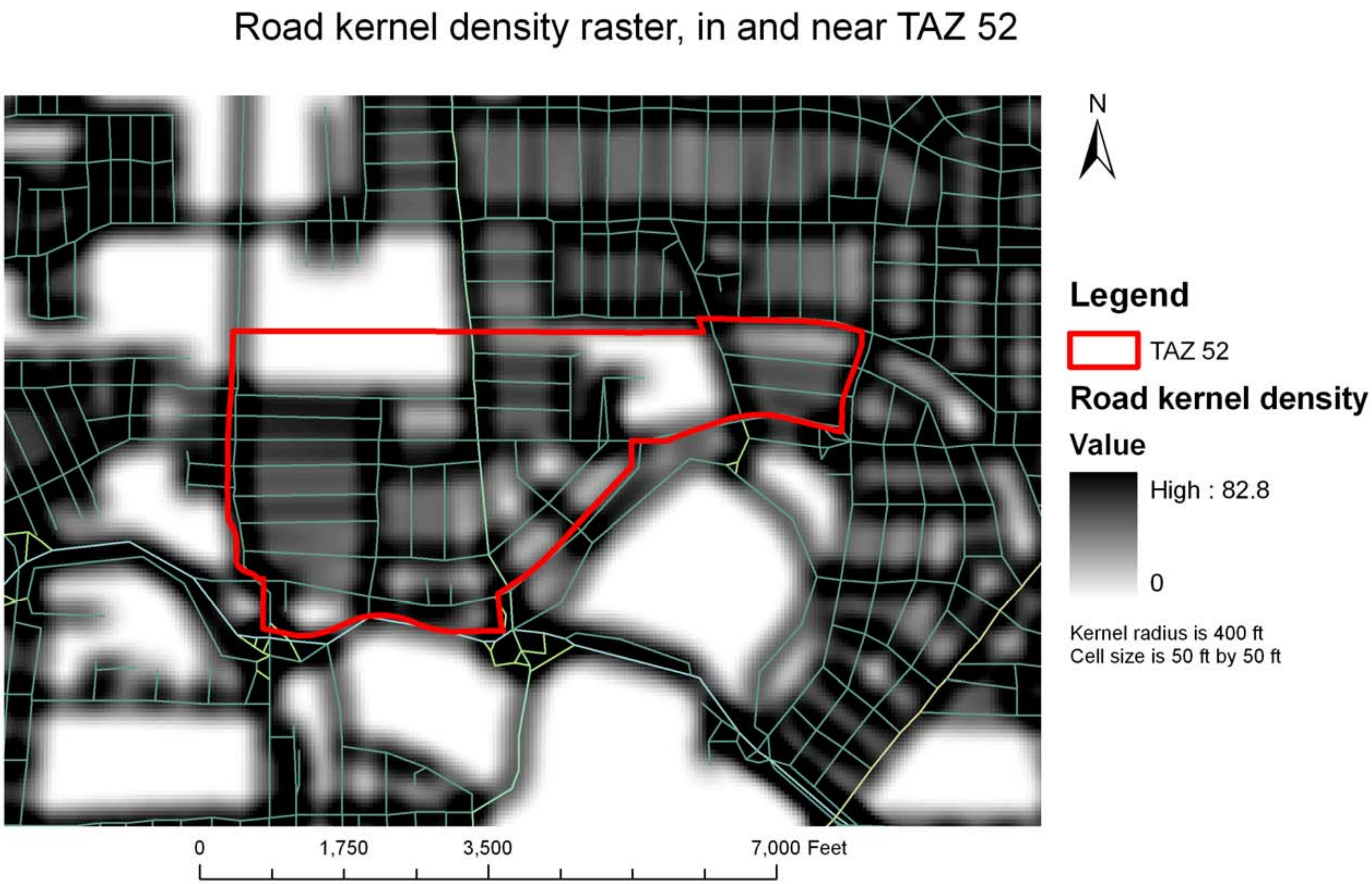


Figure C3: Road network kernel density raster map

More bla goes here

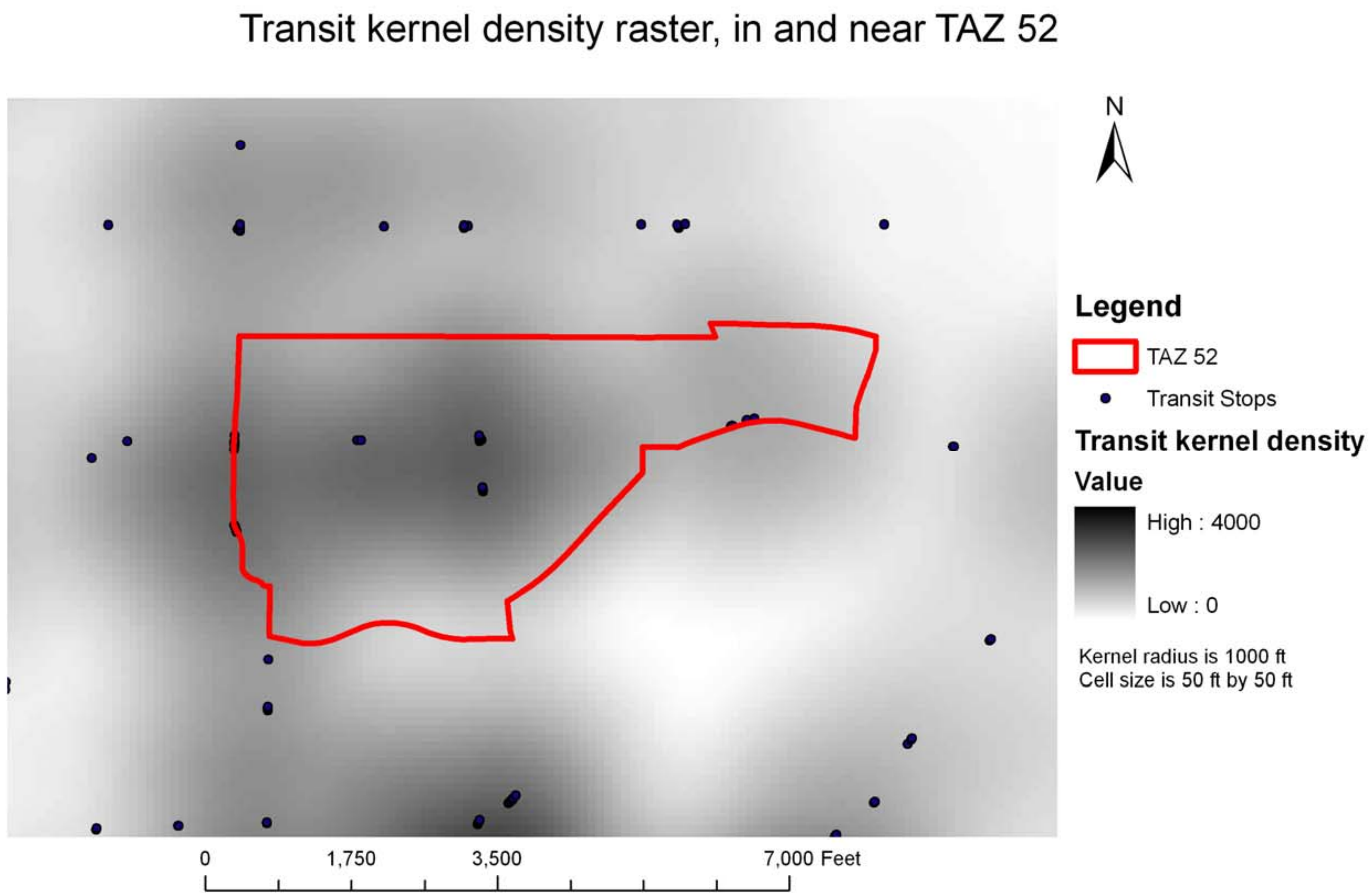


Figure C4: Transit kernel density raster map

More bla goes here

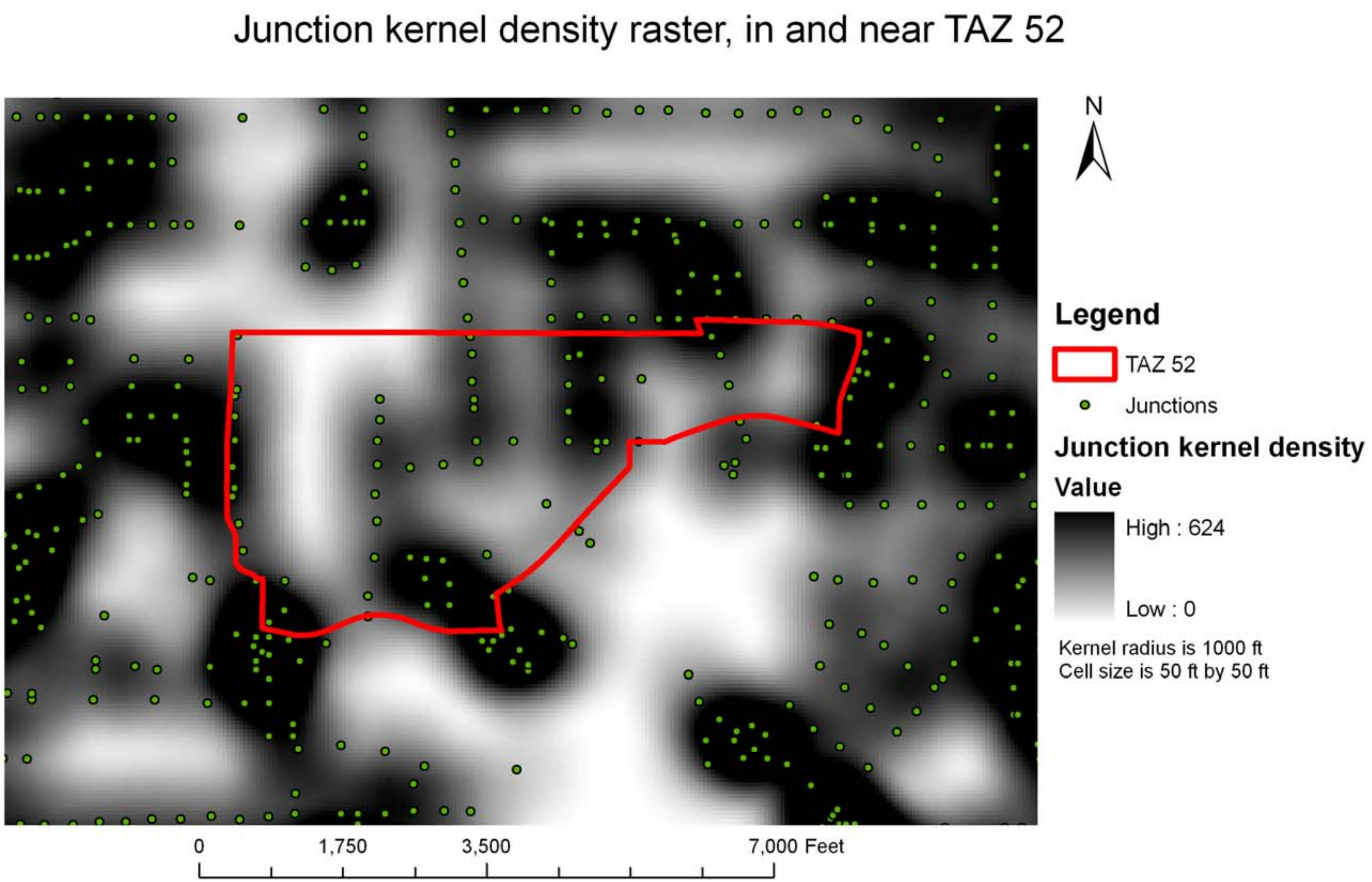


Figure C5: Junction kernel density raster map

More bla goes here

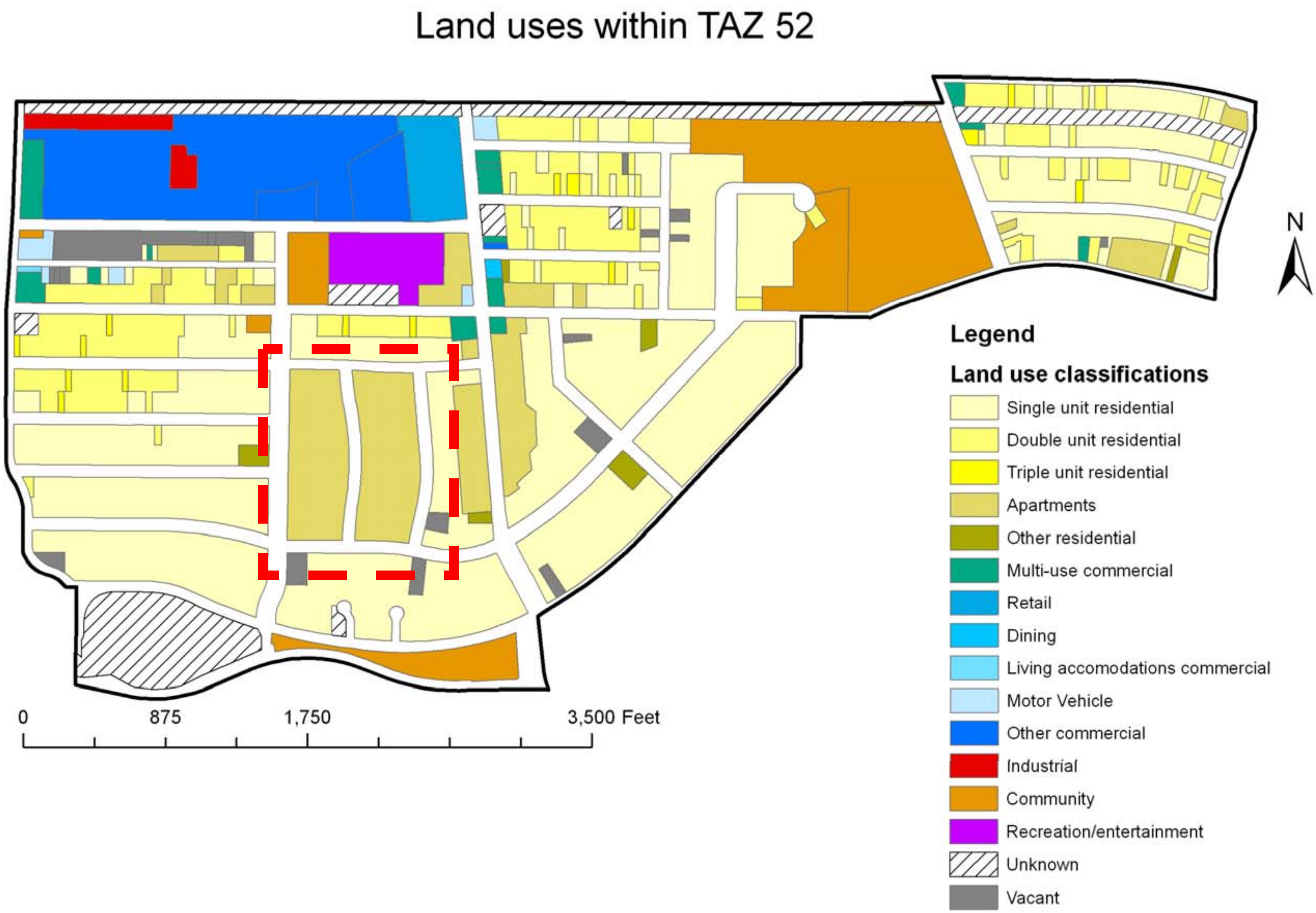


Figure C6: Land uses in TAZ 52, hypothetical land use scenario

The red box denotes the two blocks that are re-developed from low-density single unit residential housing to apartments.

Appendix D

Appendix D
Annotated R Code

```
#Initialization:

#setwd("C:/Documents and Settings/Andrew Tracy/Desktop/Andrew's
Research Shortcuts")
#setwd("K:/Research Backup")
setwd("E:/Research Backup")
ind=read.table("indfinal.txt",header=T)
dep=read.table("depfinal.txt",header=T)
indYJ=read.table("indfinalwithYJ.txt",header=T)
attach(dep)
dim(dep);dim(ind)
names(dep); names(ind)
library(leaps)
library(faraway)
library(QuantPsyc)
library(alr3)

#Summary statistics:

format(summary(dep),justify="none");
format(summary(ind),justify="none")     #for min, max, quantiles
mean(dep); mean(ind)
sd(dep); sd(ind)

#Stepwise regression:

library(leaps)
reg1adjr=leaps(ind,NonMSplit,method="adjr2",int=TRUE);
reg1cp=leaps(ind,NonMSplit,method="Cp",int=TRUE)
reg2adjr=leaps(ind,TransitSplit,method="adjr2",int=TRUE);
reg2cp=leaps(ind,TransitSplit,method="Cp",int=TRUE)
reg3adjr=leaps(ind,VehSplit,method="adjr2",int=TRUE);
reg3cp=leaps(ind,VehSplit,method="Cp",int=TRUE)
reg4adjr=leaps(ind,HB_VHT_pHH,method="adjr2",int=TRUE);
reg4cp=leaps(ind,HB_VHT_pHH,method="Cp",int=TRUE)
reg5adjr=leaps(ind,HB_VMT_pHH,method="adjr2",int=TRUE);
reg5cp=leaps(ind,HB_VMT_pHH,method="Cp",int=TRUE)

plot(reg1cp$size,reg1cp$Cp)        #used to gauge usefullness
plot(reg2cp$size,reg2cp$Cp)        #of Mallow's Cp as a stepwise
plot(reg3cp$size,reg3cp$Cp)        #regression stopping point
plot(reg4cp$size,reg4cp$Cp)
plot(reg5cp$size,reg5cp$Cp)
plot(reg1adjr$size,reg1adjr$adjr2)
plot(reg2adjr$size,reg2adjr$adjr2)
plot(reg3adjr$size,reg3adjr$adjr2)
plot(reg4adjr$size,reg4adjr$adjr2)
```

```
plot(reg5adjr$size,reg5adjr$adjr2)

plot(reg1cp$size,reg1cp$Cp,xlab="Variables in
model",ylab="Cp",main="Number of variables in model vs. Cp")
plot(reg1adjr$size,reg1adjr$adjr2,xlab="Variables in
model",ylab="Adjusted R-squared",main="Number of variables in model
vs. Adjusted R-squared")

library(faraway)
maxadjr(reg1adjr,best=5)   #retrieves five models with the best
adjusted r-squared values
maxadjr(reg2adjr,best=5)
maxadjr(reg3adjr,best=5)
maxadjr(reg4adjr,best=5)
maxadjr(reg5adjr,best=5)
reg1adjr$adjr2; reg2adjr$adjr2; reg3adjr$adjr2; reg4adjr$adjr2;
reg5adjr$adjr2   #shows stepwise adjusted r-squared values

reg1cp$Cp; min(reg1cp$Cp)
reg1cp$which [x,]   #where x is the row of the model with the minimum
Cp, or where Cp ≈ P

attach(ind)
reg1=lm(dep$NonMSplit~HHVEH+HSTUD+PopDensity+EmpDensity+ApartmentOther
ResBalance+TransitKernel+JunctionKernel+Dissim100m+ResPercent+EmpPerce
nt+HighDIndex1)
reg2=lm(dep$TransitSplit~HHVEH+HSTUD+HHSIZ+SingleResOtherResBalance+Ap
artmentOtherResBalance+SNDbyParcelArea+TransitKernel+JunctionKernel+Me
anFZ1PD+Dissim100m+CommercePercent)
reg3=lm(dep$VehSplit~MHHI_2000+HHVEH+HSTUD+EmpDensity+SingleResOtherRe
sBalance+ResCommunBalance+ApartmentOtherResBalance+SNDbyParcelArea+Mea
nFZ1PD+Dissim100m+CommercePercent+EmpPercent+HighDIndex1)
reg4=lm(dep$HB_VHT_pHH~MHHI_2000+HSTUD+HHSIZ+CommunCommBalance+ResPerc
ent)
reg5=lm(dep$HB_VMT_pHH~MHHI_2000+HHVEH+HHSIZ+RoadKernel)
#the above can also be obtained with reg1$call, reg2$call, etc.
summary(reg1); summary(reg2); summary(reg3); summary(reg4);
summary(reg5)

#Standardized coefficients:

lm.beta(reg1); lm.beta(reg2); lm.beta(reg3); lm.beta(reg4);
lm.beta(reg5)

#Correlation matrices:

round(cor(dep,ind),2)                                          #Pearson,
linear correlation matrix
round(cor(dep,ind,method="spearman"),2)      #Spearman, non-linear
(monotonic function) matrix
```

```
#Principal component analysis:

allindPCA=prcomp(ind,scale=TRUE,center=TRUE)
summary(allindPCA)
round(allindPCA$sdev,3)
names(allindPCA)
round(allindPCA$rotation[,1:8],3)
round(allindPCA$center,3)
round(allindPCA$scale,3)
plot(allindPCA,xlab="Principal Components")
biplot(allindPCA, col = c("lightgray","black"),main="Biplot of PC1 and
PC2",cex=.8)

#Plots:

hist(dep$NonMSplit,xlab="Non-motorized mode choice
proportion",main="Histogram of TAZ-level non-motorized trip
choice",breaks=15)
hist(dep$TransitSplit,xlab="Transit mode choice
proportion",main="Histogram of TAZ-level transit trip
choice",breaks=15)
hist(dep$VehSplit,xlab="Vehicle mode choice
proportion",main="Histogram of TAZ-level vehicle use")
hist(dep$HB_VHT_pHH,main="Histogram of home-based VHT per
household",xlab="Total daily VHT by all household members")
hist(dep$HB_VMT_pHH,main="Histogram of home-based VMT per
household",xlab="Total daily VMT by all household members")

hist(ind$PopDensity,breaks=10,main="Histogram of population
density",xlab="Population density, persons per square mile")
hist(yj$PopDensity,breaks=10,main="Population density, after Yeo-
Johnson transformation",xlab="Population density, after transform")
hist(ind$TransitKernel,breaks=20,main="Histogram of transit kernel
density",xlab="Mean transit kernel density")
hist(yj$TransitKernel,breaks=20,main="Transit kernel density, after
Yeo-Johnson transformation",xlab="Mean transit kernel density, after
transform")




#Variable transformations:

library(alr3)
transind=bctrans1(ind, Y = NULL, start = NULL, family = "yeo.johnson")
#applies Yeo-Johnson transformation
names(transind); summary(transind)
yj=powtran(transind)         #to output the transformed variables

reg6cp=leaps(yj,HB_VHT_pHH,method="Cp",int=TRUE)    #used to analyze
VHT and VMT models
```

```
reg7cp=leaps(yj,HB_VMT_pHH,method="Cp",int=TRUE)    #when all
variables were YJ transformed
reg6cp$Cp; min(reg6cp$Cp)
reg6cp$which
reg7cp$Cp; min(reg7cp$Cp)
reg7cp$which

log10ind=log(ind+1); round(cor(logind,dep),2)       #used to compare
Yeo-Johnson
log2ind=log2(ind+1); round(cor(log2ind,dep),2)      #to simpler
transformation functions
sqrtind=sqrt(ind); round(cor(sqrtind,dep),2)
recipind=(1/(ind+1)); round(cor(recipind,dep),2)

round(cor(yj,dep),2)
round(cor(logind,dep),2)
round(cor(sqrtind,dep),2)
round(cor(recipind,dep),2)

qq=leaps(yj,NonMSplit,method="adjr2",int=TRUE);maxadjr(qq,best=1)
#used to produce Table 6.1
qq=leaps(yj,TransitSplit,method="adjr2",int=TRUE);maxadjr(qq,best=1)
qq=leaps(yj,VehSplit,method="adjr2",int=TRUE);maxadjr(qq,best=1)
qq=leaps(yj,HB_VHT_pHH,method="adjr2",int=TRUE);maxadjr(qq,best=1)
qq=leaps(yj,HB_VMT_pHH,method="adjr2",int=TRUE);maxadjr(qq,best=1)

qq=leaps(ind,NonMSplit,method="adjr2",int=TRUE);maxadjr(qq,best=1)
qq=leaps(ind,TransitSplit,method="adjr2",int=TRUE);maxadjr(qq,best=1)
qq=leaps(ind,VehSplit,method="adjr2",int=TRUE);maxadjr(qq,best=1)
qq=leaps(ind,HB_VHT_pHH,method="adjr2",int=TRUE);maxadjr(qq,best=1)
qq=leaps(ind,HB_VMT_pHH,method="adjr2",int=TRUE);maxadjr(qq,best=1)

reg8cp=leaps(indYJ,NonMSplit,method="Cp",int=TRUE)      #regression
with selected
reg9cp=leaps(indYJ,TransitSplit,method="Cp",int=TRUE)  #variables
transformed
reg10cp=leaps(indYJ,VehSplit,method="Cp",int=TRUE)
reg11cp=leaps(indYJ,HB_VHT_pHH,method="Cp",int=TRUE)
reg12cp=leaps(indYJ,HB_VMT_pHH,method="Cp",int=TRUE)
reg8cp$Cp; min(reg8cp$Cp)
reg8cp$which [x,]            #where x is the row with the minimum Cp
reg9cp$Cp; min(reg9cp$Cp)
reg9cp$which [x,]
reg10cp$Cp; min(reg10cp$Cp)
reg10cp$which [x,]
reg11cp$Cp; min(reg11cp$Cp)
reg11cp$which [x,]
reg12cp$Cp; min(reg12cp$Cp)
reg12cp$which [x,]

attach(ind)
```

```
reg8=lm(dep$NonMSplit~HHVEH+HWORK+HHSIZ+PopDensity+EmpDensity.0.04+Res
CommBalance+SNDbyParcelArea+MeanFZ1PD.minus1.12+Dissim100m+ResPercent)
reg9=lm(dep$TransitSplit~HHVEH+HHSIZ+SingleResOtherResBalance+Apartmen
tOtherResBalance+SNDbyParcelArea+JunctionKernel+MeanFZ1PD.minus1.12+Di
ssim100m+CommercePercent)
reg10=lm(dep$VehSplit~MHHI_2000+HHVEH+HHSIZ+EmpDensity.0.04+ResCommunB
alance+ApartmentOtherResBalance+SNDbyParcelArea+MeanFZ1PD.minus1.12+Di
ssim100m+ResPercent+CommercePercent+EmpPercent+HighDIndex1)
reg11=lm(dep$HB_VHT_pHH~HSTUD.minus0.64+HHSIZ+EmpDensity.0.04)
reg12=lm(dep$HB_VMT_pHH~MHHI_2000+HHVEH+HSTUD.minus0.64+HHSIZ+EmpDensi
ty.0.04+SNDbyParcelArea+RoadKernel)
summary(reg8);summary(reg9);summary(reg10);summary(reg11);summary(reg1
2)
lm.beta(reg8); lm.beta(reg9); lm.beta(reg10); lm.beta(reg11);
lm.beta(reg12)

# Land use planning scenario

reg1$fitted.values[42]   #Where [42] is the row corresponding
reg2$fitted.values[42]   #to the scenario TAZ
reg3$fitted.values[42]
reg4$fitted.values[42]
reg5$fitted.values[42]

ind[42,]                 #Outputs explanatory variable values
dep[42,]                 #Outputs travel behavior values

scenarioA=read.table("scenarioA.txt",header=T)
scenarioB=read.table("scenarioB.txt",header=T)

predict.lm(reg1,scenarioA)
predict.lm(reg2,scenarioA)
predict.lm(reg3,scenarioA)
predict.lm(reg4,scenarioA)
predict.lm(reg5,scenarioA)

predict.lm(reg1,scenarioB)
predict.lm(reg2,scenarioB)
predict.lm(reg3,scenarioB)
predict.lm(reg4,scenarioB)
predict.lm(reg5,scenarioB)
```

Appendix E

Summary Statistics of Trip Production Data

TAZ data	Mean	Median	Standard Variation
HBW trips production (Weighted)	1142	770	1332
HBShop trips production (Weighted)	705	364	907
HBSR trips production (Weighted)	775	339	1051
HBO trips production (Weighted)	2140	1194	2628
NHBW trips production (Weighted)	368	197	476
NHBO trips production (Weighted)	1950	1097	2364
Total number of workers	1168	788	1134