MEASURING ACCESSIBILITY FOR RESIDENTIAL LOCATION CHOICE: BEYOND THE DICHOTOMY OF LOCAL AND REGIONAL

by

Xuan Liu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Urban and Regional Planning)
in The University of Michigan
2012

Doctoral Committee:

Professor Jonathan C. Levine, Chair
Professor Daniel G. Brown
Emeritus Professor Robert W. Marans
Associate Professor Joseph D. Grengs

DEDICATION

To my parents, Rongxun Liu and Lanbi Bai,

and to my wife, Yisun Cheng,

and to my daughter, Sophie

ACKNOWLEDGMENTS

This dissertation could not have existed without the help of many generous individuals. I am heartily thankful to my supervisor and dissertation committee chair, Jonathan Levine, whose guidance, encouragement, and support from the beginning of my doctoral study to the process of writing the dissertation enabled me to complete my Ph.D. degree. I would like to address special thanks to each member of my dissertation committee, Daniel Brown, Joseph Grengs, and Robert Marans who provided insightful advice and needed input at all stages of the research. I would like to thank John Nystuen who provided enormous support during my early years at the doctoral program. I also would like to express my gratitude to the current and past doctoral program chairs, Scott Campbell and Margaret Dewar for their support. I am also thankful to my fellow doctoral students, particularly Qingyun Shen for her help on modeling issues.

SEMCOG, Southeast Michigan Council of Governments, provided most of the data needed for this research. I greatly appreciate the help from SEMCOG as an organization and from all my colleagues at SEMCOG for their support. I would like to thank the following current and former SEMCOG staff particularly: Guangyu Li, Sirisha Uppalapati, Brian Parthum, Jeff Nutting, Janet Mocadlo, Delores Muller, Andy Cain, Peter McNally, Martina Nimser, Liyang Feng and Jilan Chen. I would also like to thank my predecessor, Jim Rogers, manager of Data Center, SEMCOG executive director Paul Tait and deputy director Kathleen Lomako for their unwavering support. There are many more names to mention. I am indebted to all of them, although I am solely responsible for the interpretations and any errors of omission or commission in this study.

I am also grateful to the UrbanSim team at University of California, Berkeley, and University of Washington, Seattle, particularly Paul Waddell for his guidance on SEMCOG's implementation of the UrbanSim model, Liming Wang for his support on UrbanSim programming, and Hana Sevcikova for her help on statistical matters.

My wife, Yisun, has always supported and encouraged me throughout the entire doctoral study process. I am greatly indebted to her and my daughter, Sophie, for their patience and support in this endeavor. Their contribution is immeasurable, from Yisun's "Don't worry about anything else." to Sophie's "Are you going to study tonight?" They helped me to balance academic research, professional work, and family life. I am also grateful to my parents' support and encouragement from the other side of the globe. This dissertation is dedicated to all of them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Travel demand forecasting has been a key component of long range planning at Metropolitan Planning Organizations (MPOs) in the United States. Research advancements have led to incorporating transportation accessibility into household and business location choice analysis and forecasting. The dynamic feedback effects between transportation and land use have been studied using accessibility measures with mixed results.

This dissertation examines multiple aspects of accessibility and their effects on residential location choice. First, while accessibility has been dichotomized into local and regional accessibility, this study suggests that a mid-range accessibility may have an independent and statistically significant effect on residential location choice. Second, accessibility metrics have traditionally been indifferent to the clustering of destinations. This dissertation tests the idea that, in addition to amount of activities, clustering of activities also contributes to accessibility. Models that explicitly incorporate clustering into measures of accessibility may show stronger explanatory power in predicting residential location choice than models that do not incorporate clustering into accessibility measures. Third, this study compares the effects of place-based accessibility measures and personal commute time on residential location choice. Finally, this dissertation develops alternative models for analyzing residential location choice in regard to accessibility for various socio-economic groups of population, particularly by race and ethnicity, as well as alternative models at three levels of geographic scale, which are metropolitan region, county, and city, for assessing the effects of scale on accessibility.

Research hypotheses in this dissertation are tested using multinomial logit models estimated for Detroit metropolitan area based on data from 2004 to 2010. The results show that local, mid-range, and regional accessibility measures affect residential location choice significantly, while the effects of clustering need further study. Individual workers' commute time has the biggest impact on residential location choice. This is found to be true at multiple geographic levels in Detroit region. The purpose of the study is to contribute to understanding the effects of accessibility in residential location choice, developing innovative tools for measuring accessibility that incorporate clustering and at multiple geographic scales, improving land use and transportation modeling practice, and eventually helping development of land use and transportation policies.

CHAPTER I

INTRODUCTION

Accessibility has been defined in a number of ways such as “the potential of opportunities for interaction” (Hansen 1959), “the ease with which any land-use activity can be reached from a location using a particular transport system” (Dalvi and Martin 1976), “the freedom of individuals to decide whether or not to participate in different activities” (Burns 1979), “the benefits provided by a transportation/land-use system” (Ben-Akiva and Lerman 1979), and “the ease of reaching places” (Cervero 1996). In general, the concept of accessibility concerns with how easily that various destinations in an area can be reached.

Accessibility is a driving force in models of land-use change and location choices. In general, the more accessible an area is to the various destinations in a region, the more desirable it is for people and businesses to locate, everything else being equal (Hansen 1959). The relationship between transportation accessibility and other housing and location characteristics was formalized by Alonso (1964) in a monocentric regional model. Alonso developed the economic `bid-rent' model based on the concept that residential location choices of individuals are made from a trade-off between the increasing costs of commuting to work and the decreasing land prices and housing costs when moving away from the regional employment center, typically the Central Business District (CBD). More recently, while the assumptions of monocentric regions, single-worker households, and dominating journey-to-work trips became more and more questionable, residential location choice models have been developed to account for such complexities of polycentric regions and often in the form of discrete choice models based on the random-utility theory (Mcfadden 1978; Ben-Akiva and Bowman 1998). The original feature of the classical urban economic models that treat residential location choice as a trade-off between transportation and housing cost has been generalized to account not only for travel time to the CBD but also for accessibility to a variety of destinations at regional and local scales. Accessibility is considered as one of the independent variables, along with a number of other household characteristics and place attributes, for predicting the likelihood of a household choosing a residential location. A household chooses a particular location that maximizes its utility compared to other properties, and accessibility is one important factor considered in a household's utility maximization process.

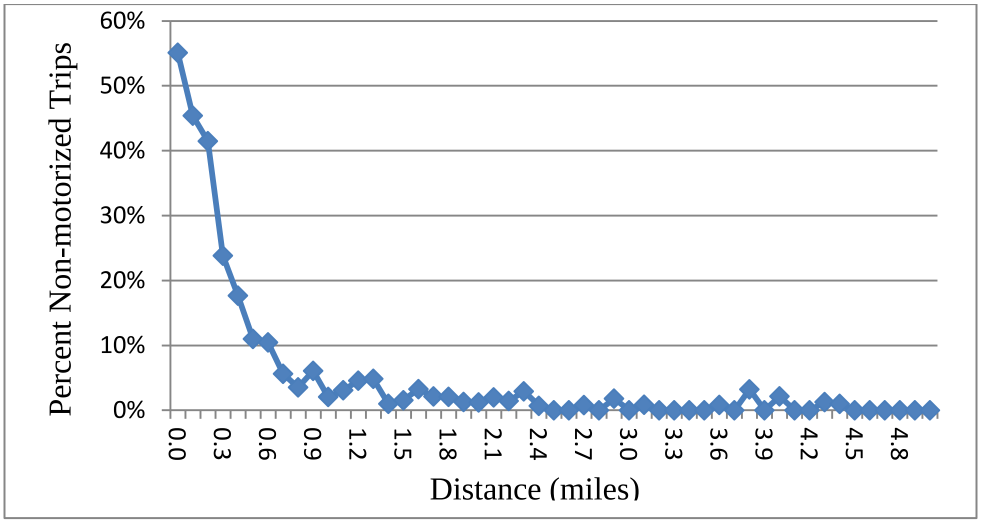
A. Scales of Accessibility

Accessibility has been defined in the regional and local ways (Handy 1993). Both regional and local accessibility were found significant in location choice models (Waddell and Nourzad 2002, Waddell and Ulfarsson 2003). "Local accessibility" typically measures access to activities by non-motorized mode (e.g. walking and biking) within a neighborhood. It focuses on neighborhood density, mix of land use, and urban design characteristics to determine how easily to reach locally oriented destinations such as drug stores, barber shops, and fitness facilities. "Regional accessibility" is a macro-scale measurement centered on the overall structure of a metropolitan region and on potential interactions within the entire region. Examples of regional destinations include employment centers (CBDs and other job clusters), airports, regional malls, and sports venues.

The primary reason for separating "local accessibility" from "regional accessibility" is to account for the difference between motorized and non-motorized trips, namely the "mode of transportation." As shown in Figure 1, the percentage of trips made by non-motorized mode declines sharply when the length of trips increases from zero to approximately one mile. The geographic range of non-motorized trips is very limited, but they provide desirable local accessibility to people who do not use motor vehicles. Local accessibility is often found significant in residential location choice models particularly for younger and smaller households.

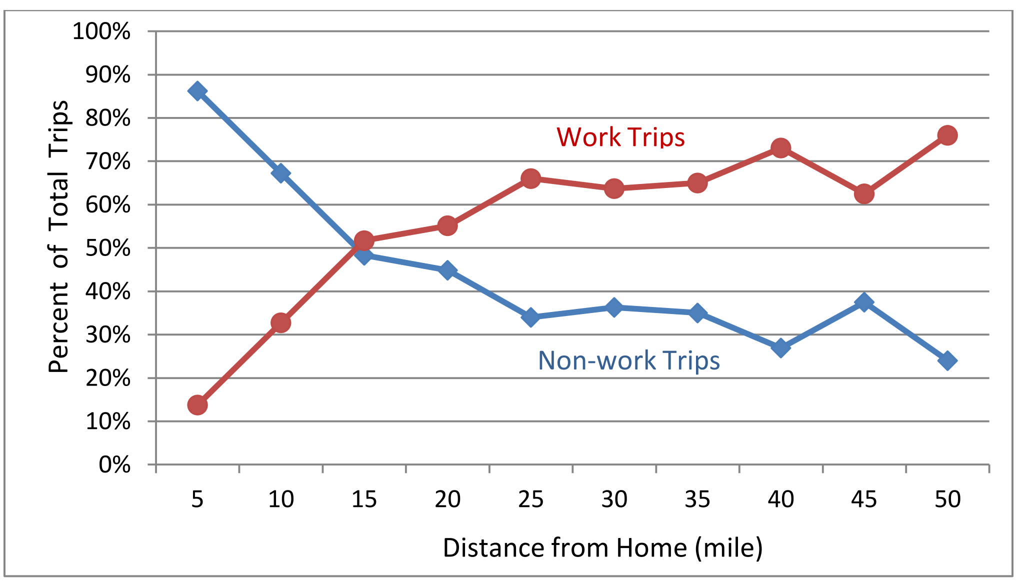
Regional accessibility could differ for different modes, different population groups, and different trip purposes. Figure 2 shows the change of work trips and nonwork trips (e.g. shopping, school, recreation) as the percentage of total trips made from home. The percent of nonwork trips declines from nearly 90% while the length of trip is five miles or shorter to about only 30% when the distance increases to 25 miles. Meanwhile work trips increase sharply. There are more work trips than nonwork trips when distance of trips reaches 15 miles. It indicates that people, in general, are more willing to travel longer distance for work than for nonwork purposes.

Figure 1. Percent Non-Motorized Mode by Trip Distance, Home-based Trips



Data Source: Southeast Michigan Council of Governments (SEMCOG) Household Travel Survey, 2004-2005

Figure 2. Percent Home-based Work and Nonwork Trips by Trip Distance



Data Source: Southeast Michigan Council of Governments (SEMCOG) Household Travel Survey, 2004-2005

The concept of distance-decay, used widely in spatial interaction models including many transportation forecasting models, can be interpreted as measuring the impedance (or unwillingness) to travel various distances to access opportunities. Evidence has shown that the degree of distance-decay varies by travel mode and travel purpose. The magnitude of empirically estimated distance-decay parameters for nonwork trips could be as much as ten times higher than for work trips (Iacono et al. 2008, p. B-7) indicating strong unwillingness to travel long distance for nonwork trips than for work trips.

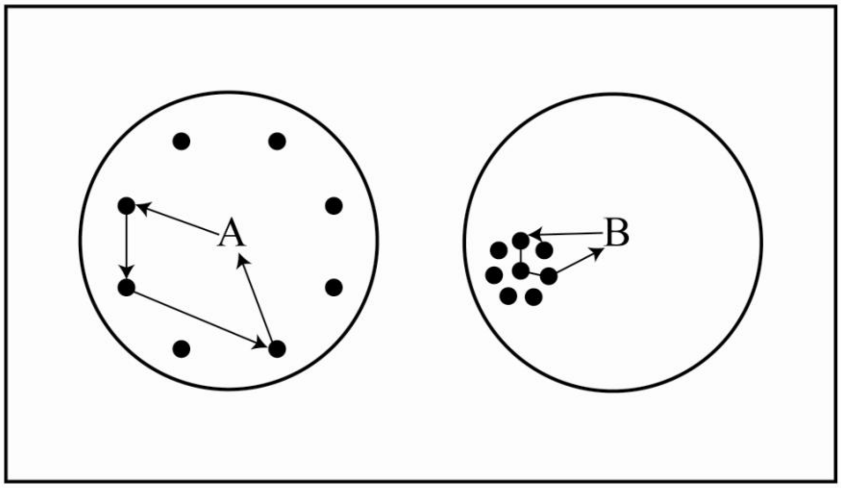
Just like local accessibility is significant for measuring the impacts of non-motorized accessibility on location choice in addition to auto accessibility, nonwork accessibility may play a different and independent role in residential location choice than work accessibility. While regional accessibility is better suited for measuring accessibility to work and regional centers, one may argue that a sub-regional “mid-range” accessibility might be needed for measuring impacts of nonwork accessibility that meets residents’ daily needs. In a metropolitan area, a household’s daily needs are mostly met at the sub-regional level. A sub-regional market provides most opportunities to its residents in regard to retail, education, and other services. When making residential location decisions, households may choose a sub-regional market area first before choosing a particular neighborhood and a specific house. It seems to suggest the importance of a sub-regional “mid-range” accessibility that is typically beyond the non-motorized range of local accessibility measures but doesn’t necessarily include all destinations in the entire region measured by regional accessibility.

B. Effects of Clustered Destinations

Since accessibility is about the ease of reaching destinations, it is conventionally measured by the amount of destinations and the cost to reach them. The simplest but widely used accessibility measure is “cumulative opportunities” method. It is simply a count of activities that can be reached within a specified threshold radius of a distance or travel time. It is most often used for measuring local accessibility where distance or travel time is short, but could be used for measuring regional accessibility as well. For example, a cumulative opportunity measure of accessibility to employment could be simply by counting “number of jobs within 30 minutes of commute.” Within this framework, higher number of jobs indicates greater accessibility. An implicit presumption in such accessibility measures is that destinations’ value is unaffected by their proximity to other destinations. As illustrated in Figure 3, both area A and B have the same number of destinations (eight in each), therefore they have the same accessibility based on conventional “cumulative opportunities” measure.

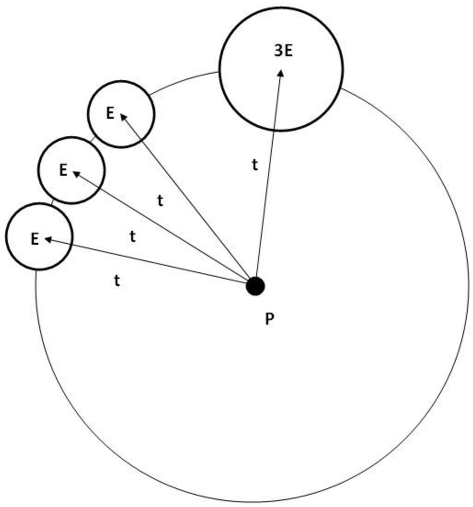
However, this is inconsistent with what we know about the travel phenomenon called “trip chaining.” If a trip has more than one destination (i.e. “trip chaining”), as shown in both A and B in Figure 3, one may think that B provides greater accessibility because the clustering of destinations in B could reduce the cost of chained trips to these destinations.

Figure 3. Clustering Effects in the Cumulative Opportunity Measure



The effects of clustering apply to other conventional accessibility measures as well. For example, the gravity model is widely used in measuring accessibility. While the "cumulative opportunities" method treats all destinations within the fixed threshold equally by omitting distance or travel costs and ignores all destinations beyond the fixed threshold, a gravity model overcomes those weaknesses by including all destinations in the region and by incorporating travel costs to recognize that distant destinations are less desirable than close destinations. But it also treats destinations equally regardless whether they are in the clustered urban centers or at the dispersed rural periphery of a region. As illustrated in Figure 4, based on conventional gravity model, accessibility from location "P" to the larger destination "3E" is the same as the sum of accessibility from "P" to the three smaller destinations, as long as "t" is the same impedance for all destinations and the attractions at the larger destination, 3E, is three times the attractions at each smaller destination E. The "attractiveness" in conventional gravity-based accessibility measures is often measured by the *amount* of activities alone, such as "number of jobs." However, there are other factors besides the amount that can affect the attractiveness. Clustering is one of them.

Figure 4. Clustering Effects in Gravity-Based Regional Accessibility



One may argue that destination “3E” is more important than the sum of the three “E”s. It is likely that “3E” has greater regional impacts, i.e., attractiveness to the entire region, for its being able to provide higher level of services to the region. The hierarchical central place theory (Christaller 1933/1966) asserted that settlements functioned as 'central places' providing services to surrounding areas. A region is consisting of nested market areas of various economic scales. The higher the order of the goods and services, the larger the range of the services, the longer the distance people are willing to travel to acquire them. Destination “3E” is likely to be at a higher level of the regional hierarchy than the three “E”s. Access to “3E” is more valuable than to the “E”s. This could also be in part because of the clustering of destinations that makes trip chaining more efficient. If a trip from “P” has more than one stop, destination “3E” has the potential to provide better accessibility for the intervening stops by clustering stops than the more scattered three “E”s. In summary, when destinations are clustered, the whole could be greater than simply the sum of its constituent parts. Therefore, it could be desirable to weight destinations by degree of clustering when estimating accessibility, regardless which accessibility measuring method to use.

Trip chaining is a common phenomenon in daily travel. Analysis of Detroit region household travel survey shows that 43% of home-based tours (a round-trip journey from home) had more than one destination, i.e. chained tours. More importantly, 70% of all destinations in the survey were reached by chained tours (Table 1).

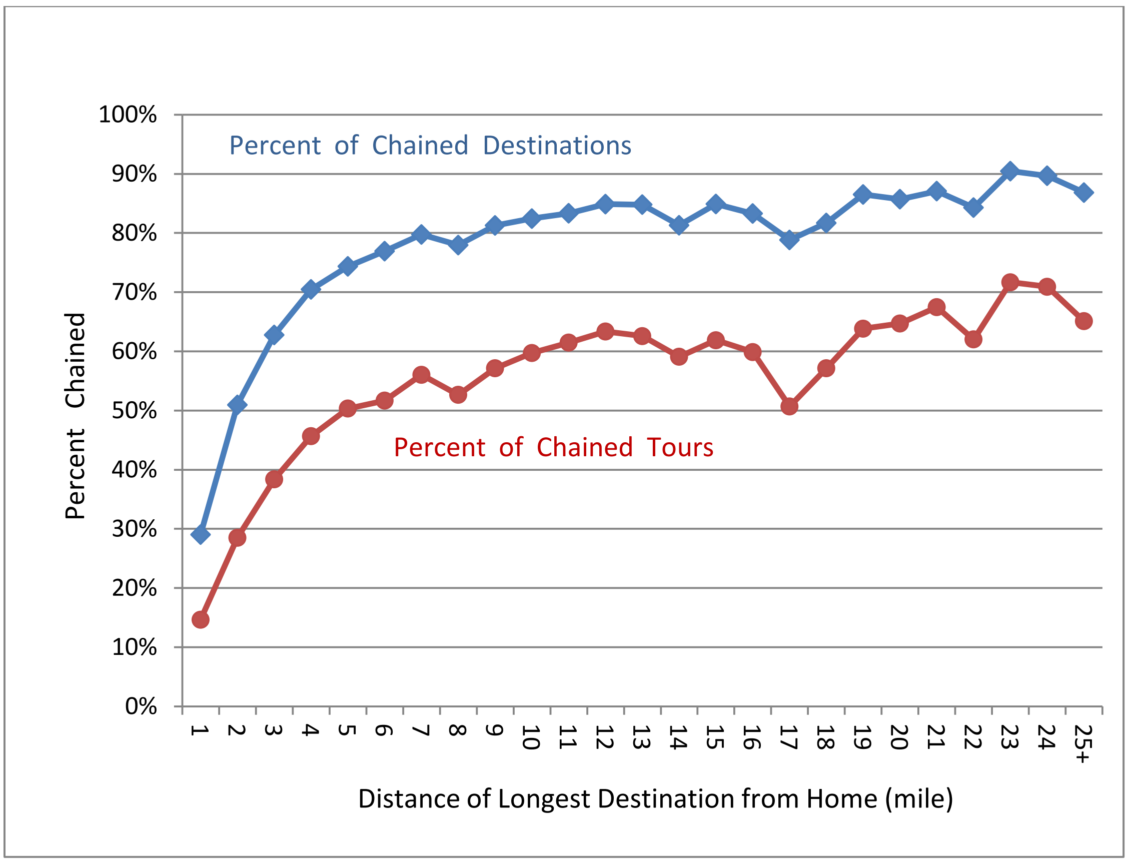
If we assume that clustering carries no weight in accessibility measures, we are effectively assuming that interaction potential is always measured from home with one destination, i.e. home-destination-home. Under that assumption, clustering would not matter. However, since there is high potential for interaction that goes home-destination1-destination2- … destination-n … -home, clustering would be essential because it reduces the impedance between intervening destinations.

Table 1. Chained and Unchained Trips, based on Detroit Region Household Travel Survey, 2004-2005

	Chained	Unchained	Total	Percent Chained
Tours	4,861	6,418	11,279	43%
Destinations	14,702	6,418	21,120	70%

More detailed analysis of the same travel survey data shows that the likelihood of forming chained trips increases when trip length increases (Figure 5). When distance of the longest destination from home reaches five miles, more tours are chained (51%) than not. Furthermore, 82% of destinations are reached by chained tours when distance increases to ten miles. While intervening stops could be either along the way to the longest destination or clustered around the "final destination", it suggests clustering of destinations may have significant impacts on regional accessibility.

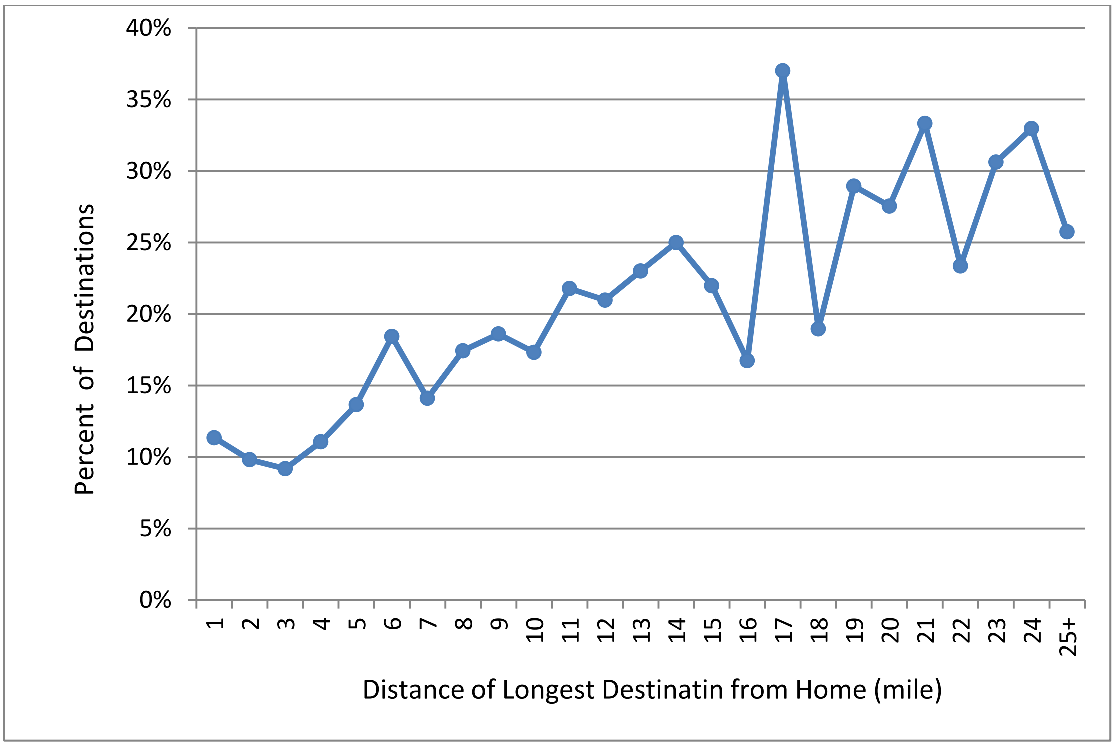
Figure 5. Percent Chained Tours, Destinations, and Distance from Home



Data Source: Southeast Michigan Council of Governments (SEMCOG) Household Travel Survey, 2004-2005

Further analysis seems to show that destinations increasingly cluster to the end of tour when distance of travel increases (Figure 6). Overall, clustering seems to have more impact on regional accessibility than local accessibility. It suggests a form of regional accessibility that considers a destination's proximity to other destinations in measuring its effects and assessing its importance in location choice.

Figure 6. Percent of Destinations with Distance to the Longest Destination less than One Fifth the Distance to Home in Chained Tours



Source: Southeast Michigan Council of Governments, SEMCOG, Household Travel Survey, 2004-2005

In summary, the conventional approach of dichotomizing regional and local accessibility seems to neglect the potential influence of an independent mid-range accessibility. Furthermore, the omission of clustering in regional accessibility seems to neglect the importance of trip chaining, particularly trip chaining where stops are closer to the far end of the trip than they are to home.

C. Commute Time

The above sections have discussed place-based accessibility measures that focus on the spatial separation and connection among locations such as home, employment centers, retail stores, health care services, and recreation facilities. There are other implementation issues associated with these place-based accessibility measures besides the scale issues and clustering issues that are discussed in Section A and B respectively. One issue is related to aggregation. Accessibility measures typically use some kind of spatial zones to group activities. For example, accessibility to employment measures often use number of jobs in Traffic Analysis Zones (TAZs) to represent job attractiveness to all workers in a region. However, geographic aggregation of destinations can affect the results of accessibility analysis. If the zone system changes, the measured results of accessibility may differ. Existing spatial analysis literature defines this issue as Modifiable Areal Unit Problem (MAUP), that is if the zone system is modifiable, the results can vary simply because of changing the zone system (Cressie 1996). The problem is twofold. First, the level of spatial aggregation is artificial. Second, the delineation of the zones is also artificial. There is no perfect solution to this problem. A general suggestion is spatial disaggregation, and greater disaggregation is better (Handy and Niemeier 1997). Miller (2005) believes that place-based measures of accessibility should be enhanced and complemented with people-based measures that are more sensitive to individual activity patterns and accessibility in space and time. He argues that only complete disaggregation to the atomic units of analyses can eliminate MAUP problem.

When analyzing accessibility to employment, the "atomic unit of analysis" is an individual worker's journey to his or her own job. A worker's specific employment location may affect his or her residential location choice most. When there are multiple workers in a household, multiple employment locations of these workers may all affect the household's residential location decision. However, this is not to suggest that individual workers' commute time should replace place-based accessibility. A household may choose to locate near its workers' employment locations. It may also consider place-based overall employment accessibility as well for potential employment changes of its workers in the future. Places with good overall employment accessibility maybe even more important for households with multiple workers than single-worker households, because high overall accessibility of a place may help optimize individual employment accessibility for all workers in the household.

D. Research Objectives

The effect of accessibility on residential location choice is a controversial research topic. There is contradictory empirical evidence to support or dismiss such effects. This study examines the scale issue and the clustering issue of the accessibility concept and its measurement by exploring the role of accessibility in residential location decisions. How effectively can residential location choice be explained by a range of accessibility from regional, to sub-regional, and to local, and by incorporating clusters of destinations? Do place-based accessibility and individual workers' commute time affect residential location choice differently? The study examines the notion that accessibility at each scale has its own power in explaining where households choose to live, adding clustering effects improves the explanatory power, and both place-based accessibility and workers' commute time affect residential location choice.

Specifically, the hypotheses of the study are as follows:

First, a sub-regional mid-range accessibility will have a statistically significant effect on residential location choice, controlling for regional and local accessibility.

Second, models that incorporate clustering into measures of accessibility will show stronger explanatory power in predicting residential location choice than models that do not incorporate clustering into accessibility measures.

Third, both place-based accessibility and workers' commute time affect residential location choice significantly, although their effects may not be equal.

These hypotheses are tested using multinomial logit models estimated for Detroit metropolitan area based on data from 2004 to 2010. Alternative models for various population groups with different socio-economic characteristics are developed to assess the different effects of accessibility on their residential location choice.

The purpose of the study is to contribute to understanding of accessibility at various scales in a region, providing additional tools for measuring accessibility characteristics of metropolitan areas that can be used for predicting household location choices, improving land use and transportation modeling, and eventually helping evaluation of land use and transportation policies.

E. Organization of Chapters

This dissertation is organized in three main parts including six chapters. The first part (Chapter I and II) discusses the theoretical perspectives on the effects of accessibility on residential location choice. Chapter I introduces the concept of “mid-range” accessibility in addition to regional and local accessibility. This chapter also argues for the effects of clustering of activities on accessibility. Chapter II reviews the literature on accessibility, its measurements, and effects on location choice. Central Place Theory is also discussed as a theoretical base for including “mid-range” scale and clusters in measuring accessibility for residential location choice.

The second part (Chapter III and IV) presents a descriptive analysis of accessibility across a region and its residents’ location choices. Chapter III provides an overview of the study area, the Detroit region in Southeast Michigan, and explores all data items that are necessary for analyzing accessibility, assessing its impact on location choice, and testing research hypotheses. These data items include demographic data, socio-economic data, and transportation data. This chapter illustrates the land use characteristics and travel patterns in the region. Chapter IV discusses the methodology used for this research. Various “mid-ranges” are tested for measuring accessibility. The degree of clustering of activities is measured to weight accessibility. Multinomial logit models are estimated for analyzing residential location choice to test the effects of “mid-range” accessibility and clustering of activities, controlling for other variables. Workers’ commute time measures are used with place-based accessibility measures in these models.

The third part (Chapter V and VI) presents the modeling results and discusses their implications. Chapter V first presents the results of models with various accessibility measures. It then discusses alternative models for various socio-economic groups of population, as well as alternative models at various scales within a region. It also interprets modeling outcomes and reviews the research hypotheses. Chapter VI discusses the implications of the findings from this study, in regard to planning theory, land use and transportation modeling practice, and enhancing public policy. It addresses the values of this research that may contribute to urban and regional planning, and suggests further research questions in the future. Finally, the dissertation concludes by highlighting the lessons learned from this research.

F. Summary of Chapter I

Accessibility is a driving force in location choices. The more accessible an area is to the various destinations in a region, the more desirable it is for people to locate there, all else being equal. The relationship between accessibility and residential location choice has been studied in various research projects. Accessibility has been dichotomized to regional and local accessibilities in some of these research projects. While "local accessibility" measures access to activities within a neighborhood scale and is thus particularly relevant to travel by non-motorized modes, "regional accessibility" centers on the overall structure of a metropolitan region and on potential interactions within the entire region. Both local and regional accessibilities have been found significant in residential location choice in previous studies.

This study tests whether a sub-regional mid-range accessibility is needed in assessing accessibility and estimating the effects of accessibility on residential location choice. While the primary reason for separating “local accessibility” from “regional accessibility” is to account for the difference between non-motorized and motorized trips, “mid-range” accessibility may identify the uniqueness of nonwork accessibility from work accessibility. This mid-range accessibility may play an independent role in residential location choice, in addition to local and regional accessibility.

This study also suggests that the clustering of activities, besides the amount of activities, is important in analyzing accessibility. If we assume that clustering carries no weight in accessibility measures, we are effectively assuming that interaction potential is always measured from home with one destination. Under that assumption, clustering would not matter. But since there is high potential for interaction that involves multiple destinations as demonstrated in trip-chaining, clustering would be essential because it reduces the impedance between intervening destinations. Models incorporating clustering into measures of accessibility will show stronger explanatory power in predicting residential location choice than models that do not incorporate clustering into accessibility measures.

Furthermore, this study argues that both place-based accessibility and individual workers’ commute time affect residential location choice. They may co-exist as independent variables in residential location choice models. Including commute time to work in residential location choice model may supplement other independent variables that measure place-based accessibility.

CHAPTER II

LITERATURE REVIEW: APPROACHES TO UNDERSTANDING ACCESSIBILITY AND RESIDENTIAL LOCATION CHOICE

A. The Concept of Accessibility

Accessibility and its effect on location decisions as a research topic have a long tradition in literature related to geography, urban economics, and planning. Hansen (1959) defined accessibility as “the potential of opportunities for interaction.” His 1959 article has been often cited in accessibility research as seminal to the concept of accessibility and constructing the measurements of accessibility. He presented an operational definition and suggested a method for determining accessibility patterns within metropolitan areas. His formulation of accessibility states that the accessibility is “directly proportional to the size of the activity … … and inversely proportional to some function of the distance” between locations. It is also important to note that Hansen’s intention was to develop a residential land use *model* that relates accessibility of an area to the rate and intensity of land development in that area “based on a realistic measurement of accessibility.” The primary focus of his research was “an empirical examination of the residential development patterns illustrates that accessibility and the availability of vacant developable land can be used as the basis of a residential land use model” (Hansen 1959, p. 73). Since then, accessibility has been proved to be a useful tool for metropolitan planning purposes. Henson’s concept of accessibility and approach for measurement have been continuously refined in the past decades, particularly in the 1990s and 2000s.

Handy (1993) summarized research advances in accessibility and categorized accessibility into local and regional types. Local accessibility is defined with respect to "convenience" establishments. Only such establishments that are nearby or that are nearest to people are included in local accessibility. These establishments usually are found in small centers or in stand-alone locations in neighborhoods. Local accessibility is presumed to be associated with short and relatively frequent "local" trips, whereas the choice of particular destinations will depend to a large degree on the distance to that destination. On the other hand, regional accessibility is defined with respect to regional employment centers, suburban shopping malls, or other major commercial areas, which offer an abundant of job opportunities, or a wide range of "comparison" of goods and services. These activity centers may be close to neighborhoods or relatively far. They attract people from a wide geographic area. Regional accessibility is associated with longer regional trips, where distance is less of a concern in destination choice compared to local accessibility. But, Handy also noted that the distinctions between regional and local accessibility “are not entirely clean” (Handy 1993, p. 59). There is no definitive geographic scale specified for local or regional accessibility.

In fact, the geographic definition of “local” and “regional” in measuring accessibility varies significantly. “Local accessibility” ranges from TAZs and “super districts” (Handy 1993) to “walking distance” (Waddell and Nourzad 2002). Increasingly, “local accessibility” has been defined as “neighborhood accessibility” (Krizek 2003) that is geographically limited to “walking distance.” Therefore, local accessibility is a micro-level measurement that typically focuses on density, land use mix, and urban design characteristics (Cervero and Kockelman 1997) at the neighborhood scale. It is particularly helpful in measuring the effects of non-motorized modes of transportation, which are often overlooked by traditional transportation planning tools. However, it left everything beyond “walking distance” to regional accessibility, regardless distance, travel mode, and trip purpose.

“Regional accessibility” is a macro-scale measurement centered on the overall structure of metropolitan regions and focused on potential interactions at the regional scale. The interaction between local and regional accessibility has not been explicitly studied but deserves more research. Accessibility of a walkable neighborhood in the middle of nowhere is very different from accessibility of a walkable neighborhood within a well established city. Under the right conditions, there may be some substitutability between local and regional accessibility. A high level of local accessibility may reduce the frequency of regional trips, whereas a high level of regional accessibility may reduce the frequency of local trips, depending upon the characteristics of these trips. For example, although it is difficult to substitute work trips because of the fixed job locations, trips to retail stores and service centers could be more easily substituted. Furthermore, it is hard to justify a long trip to a single destination. But if there are multiple destinations clustered at a particular location, people may be more willing to travel longer distance. The concept of accessibility is truly multi-dimensional. The above discussions touched on some important dimensions of accessibility, such as regional vs. local, and work vs. nonwork. The rest of this chapter will further explore these dimensions and other dimensions of accessibility, including travel mode and place-based vs. people-based accessibility.

Overall, accessibility is determined by a number of factors including: 1) the spatial distribution of activities, 2) the magnitude, quality, and characteristics of the activities, and 3) the means of reaching those activities. Destinations where activities take place are central: the more destinations with greater varieties, the higher the accessibility. Travel cost is crucial: the less time (or cost in general) spent in travel, the more destinations that can be reached within a budget, the greater the accessibility. Travel choices are important too: the wider the variety of modes for getting to destinations, the better the accessibility. Personal preferences and constraints also make significant differences. “Accessibility is thus determined both by patterns of land use and by the nature of the transportation system, although two people in the same place may evaluate their accessibility differently, as wants and tastes vary” (Handy and Niemeier 1997).

The multi-dimensional nature of accessibility makes it difficult and complex to measure. A number of ways of measuring accessibility have been proposed, although what constitutes the best or even suitable measures of accessibility is far from clear. Specifications of accessibility measures have varied substantially from simple minimum travel time indices (Leake and Huzayyin 1979), measures of cumulative opportunities within specified distance or time thresholds (Wachs and Kumagi 1973), gravity-based measures (Wilson 1971), to maximum utility functions (Niemeier 1997). These measures will be further discussed in Section C.

B. Accessibility and Residential Location Choice

The issue of accessibility's effects on residential location choice is controversial. There has been contradictory empirical evidence to support or dismiss such effects. While Hansen (1959) and others believed that accessibility is essential to explain metropolitan development patterns, and Alonso (1964) proved that residential location choices of individuals are made from a trade-off between the increasing costs of commuting to work and the decreasing land prices and housing costs when moving away from the regional employment center, others have questioned the importance of accessibility in residential location decisions.

Hamilton (1982) questioned whether commuting behavior can be predicted by the classic monocentric model developed by Alonso (1964). He instead introduced the concept of "excess commuting" to argue against the impact of accessibility on residential location choice. The excess commuting approach divides "actual commuting time" into "required minimum commuting time" and "excess commuting time." Its methodology is based on the assumption that individual households, each minimizing its housing cost and commuting cost, will achieve an equilibrium with no "cross-commuting," which is one that minimizes aggregate commuting cost given the distributions of housing and job locations. White (1988) tested this theory by applying a linear programming method to the existing distribution of housing and job locations, reassigning workers to housing locations so as to minimize average commuting cost. That is, the assignment algorithm minimizes the following quantity:

$$Z = \sum_{i} \sum_{j} C_{ij} X_{ij}$$

subject to the constraints for every *i.j.*:

$$\sum_{i} X_{ij} = D_j;\ \ \sum_{j} X_{ij} = O_i;\ \ and\ X_{ij} \geq 0$$

where *Xij* is the number of workers commuting from zone *i* to zone *j*; *Cij* is the corresponding travel cost (e.g., commuting time); *Dj* is the employment in zone *j* ; and *Oi* is the number of workers residing in zone *i*.

This is a cost-minimizing assignment approach based solely on job locations and resident locations. Under this approach, a journey-to-work matrix (origin–destination matrix) is constructed to contain the elements *Xij* showing the number of workers commuting from their residential place (zone *i*) to the work place (zone *j*). A corresponding matrix of commuting costs is also constructed to contain the elements *Cij*. The linear programming reassigns the locations of jobs and residences to minimize the total average commuting costs to find the optimal journey-to-work flow in the origin–destination matrix. It then compares the minimum commute resulted from linear programming analysis to the actual commute. Various excess commute studies have found a wide range of excess commuting (a.k.a., "wasteful commute") from 11 percent to 87 percent of actual commute (Ma and Banister 2006).

Several studies on excess commuting have tried to explain why excess commuting exists. Most pointed to the simple assumptions of the bivariate relationship between jobs and residents' locations that formed its basis. Some of the reasons that have been mentioned in those studies are as follows.

- Multi-worker households
- Tenancy
- Uncertainty of job locations

- Heterogeneous housing and job markets
- Different tax subsidy systems
- Minority groups
- Moving costs
- Neighborhood amenities
- Rapid job turnover
- Decreasing importance of work trips

(Ma and Banister 2006, p. 754)

Although these factors have been mentioned as possible explanations of excess commuting, very few of them have been incorporated into the excess commuting models. Yet, conclusions have been drawn from this over-simplified approach that job locations have only a limited influence on residential location choice and therefore accessibility is less important in shaping urban physical forms.

Some studies on excess commuting acknowledged this approach's major weakness of omitting factors other than commuting and attempted to incorporate other variables into their models. Giuliano and Small (1993) tried to add worker occupations as a 'constraint' in their cost minimization procedure. They identified seven occupational groups in the data from Los Angeles region. Adding this occupational constraint means to do the cost minimization seven times, once for each group. But there are at least three unresolved issues in this approach. First, the problem of this approach is still assuming that people can trade places freely to minimize commute, as long as they are in the same occupation, regardless if a specific job fits the commuter. Second, even if there is more than one job in an occupation that fits a particular commuter, job selection is probabilistic in realty due to other factors rather than deterministic in the excess commuting model that is only subject to minimizing commuting. Job interchangeability is less common in the real world than in those models. Third, this modified approach is still not a true multivariate analysis and ignores many other factors in the analysis. There are many other variables that need to be considered simultaneously. Nevertheless, adding occupational constraint increased “average required commute” by 22 percent. It indicates exactly that there are other factors need to be considered in those models. Commute time is only one factor. However the authors stated that “mismatches could lengthen commutes to some extent, but more than half of the average commute time remains unexplained” to conclude that “job location has only a limited influence on housing-location choice” (Giuliano and Small 1993, p. 1488).

Manning (2003) attempted to further incorporate heterogeneous household characteristics and job sub-markets by occupation into excess commuting analysis. He disaggregated London data into a number of characteristics such as age and occupation, and checked whether there is a convergence of minimum required commute towards actual commute when these restrictions are considered. Using the Greater London data from the 1991 Census, Manning analyzed seven different age categories and 23 occupation groups. He found only very small effects of disaggregation on the volume of excess commuting. In the male labor market, imposing an occupational constraint caused an increase in the minimum commute by an average of only 30 meters. Similarly, age disaggregation caused a rise in the minimum commute by only 40 meters, which in practical terms means that there is a minimal effect on the amount of excess commute. Manning argued that excess commuting is not likely to disappear by imposing more restrictions. He indicated that even among workers doing the same job, there is a large variation in pay, which is likely to lead to a certain amount of excess commuting.

The findings in Manning's study are not surprising. It actually conveys a strong message about the importance of including a full range of variables in analyzing accessibility and location choice. The number of variables, or "constraints," that can be added to excess commuting analysis is limited due to its unusual methodology. The distribution of jobs and households in the real world is never likely to be 'optimal' as defined by excess commuting approach. This approach attempts to explain spatial structure solely in terms of commuting. It is an over-simplified method to use for analyzing such a complex socio-economic issue. It needs to be extended to include a wide range of individual, social, and location factors.

In summary, despite some intuitive appeal of distinguishing 'excess commuting' from 'inevitably necessary commuting,' the excess commuting approach can only explain a part of location choice decision making. Despite variations of the methodology and several attempts for improvement, it failed to address a wide range of individual, social, and location variance in location choice. What more problematic are the conclusions drawn from excess commuting analysis. It is wrong to conclude that access to jobs is not important because commute time cannot explain *all* the variance in location choice. The excess commuting studies employ an unusual construct in its methodology. It is based on an extremely strong assumption that simplifies the reality, which is minimizing journey-to-work by considering job locations and residence locations *only*. It is basically a bivariate analysis that expects commuting itself to fully explain residential location choice decisions. It ignores many other important factors that affect commute and residential location choice in turn. However, in a more typical social science approach, one would consider as many variables as possible in analyses.

Other studies have used residential location choice models to predict households' choices of where to live based on a wide range of variables including the characteristics of households and the attributes of locations. Accessibility, used as an independent variable in residential location choice models, is a measure of how well transportation options interact with land use attributes that satisfy household preferences.

Srour et al. (2002) developed models that related general accessibility indices for the Dallas-Fort Worth region of Texas to property valuations for single-family dwelling units and commercial units, and to household location choices. Multinomial logit models were used to derive logsum measures of accessibility as well as to assess the effect of accessibility on location choices, while controlling for household demographics. They considered independent variables such as average lot size and value, home size, value and age, average number of garages and bathrooms, and distance to work. They developed four different forms of regional accessibility measures including logsum-based, cumulative opportunities, and residential land value residuals by census tract or by parcel. Four models were developed, one for each accessibility measure along with all the control variables. The results of the four model estimations were compared. They found that various functional specifications of accessibility measures appeared useful. "Cumulative opportunities access measures were most helpful in predicting residence location (Srour et al. 2002, p. 25)." This was because it had the highest t-value (5.22) among the four accessibility measures, and the model that included this measure had the highest "goodness of fit" measure ($\rho = 0.38$). Meanwhile, the t-values for control variables ranged from -0.17 (average garage) to -19.4 (work distance). While individual households are the choosers in their model, Traffic Analysis Zones (TAZs) are the units in the choice set. This limits the accessibility measures used in the model to be regional. Local accessibility was excluded from the model.

Waddell and Nourzad (2002) developed a residential location choice model in a set of land use models called UrbanSim, in which geographic unit of analysis is disaggregated to land parcels or gridcells of five acres each. Their residential location choice model included both regional and local accessibility to predict the probability that a household is likely to choose a housing unit at a parcel or gridcell. The form of the model is specified as multinomial logit. The data used in the model draw principally from a household travel survey conducted at the Puget Sound Region in the State of Washington. The independent variables in the model include household characteristics such as income, motor vehicle availability, number of workers and children. Location attributes include assessed value, density, development types, commercial, industrial, and institutional building, as well as regional and local accessibility.

They operationalize the concept of regional accessibility for a given location as the distribution of opportunities in the region weighted by the composite utility of all modes of travel to those destinations, defined as the logsum from the mode choice model for each origin-destination pair. Specifically, regional accessibility to employment opportunities in the region is represented as the sum of logsum from a TAZ to all other TAZs multiplied by employment numbers at those TAZs (same as Equation 5 shown in Section C on pages to follow). Model estimation results show positive and significant coefficients for regional accessibility to employment at each household auto ownership class (Waddell and Nourzad 2002, p. 123), which means that households are more likely to choose locations with better access to employment in the region, everything else being equal. In other words, the model estimation results indicate that households prefer to reduce commuting costs to work by automobile or by transit while controlling for other factors, because driving and taking transit are the two modes included in the impedance measurement using the logsum.

Local accessibility is also considered in the UrbanSim model by including such variables as density and mixed uses in neighborhoods. It is found that some interactions between household attributes and local accessibility are significant. For example, young households are more likely to choose high density and mixed-use neighborhoods. Overall, in the Waddell and Nourzad (2002) model, several local and regional accessibility measures are statistically significant, with t-values up to 4.64 (variable "log access to employment for two-car households"), meanwhile the t-values for control variables range from -1.57 (a dummy variable for moderate high density residential development) to -15.05 ("log total number of housing units in cell"). The model's goodness of fit measure, i.e., log likelihood ratio, was 0.13.

The UrbanSim modeling system has been implemented in a number of metropolitan areas in the United States, including the Detroit region (Waddell et al. 2008). In the Detroit region's UrbanSim system, both regional accessibility to jobs (t-value = 4.15) and local accessibility interacting with young households (t-value = 3.64) are found positive and statistically significant in the residential location choice model. Meanwhile, some other variables are more significant in the model. For example, percent of minority households within walking distance for a minority household has a very high t-value of 27.61. And the t-value for income interacting with housing value is 14.04. The overall model goodness of fit measure is 0.17.

Using conventional regional and local accessibility in predicting household location choice does not account for the clustering effects of destinations, and may have under-estimated some "mid-range" accessibility, which may have its own and independent impact on location choices. While place-based accessibility measures are often used in residential location choice models, people-based accessibility may also show significance using disaggregated data and analytical methods.

C. Accessibility Measures

Existing approaches of measuring accessibility can be classified into three broad categories: cumulative opportunities, gravity-based, and utility-based.

1. Cumulative Opportunities

A simple form of measuring accessibility is a count of activities that can be reached within a specified threshold radius of a distance or travel time, e.g., walking distance. For example, the following equation (1) measures the accessibility of location *i* to shopping opportunities by counting number of stores within walking distance using spatial queries. The more number of stores, the higher the accessibility.

$$Access_i = \sum_j Stores_j \,\forall\, WalkingDistance \qquad (1)$$

Cumulative opportunity measures estimate accessibility only in terms of the quantity of opportunities available within an arbitrary limit. The advantages of these measures are related to its operationalization, interpretability, and communicability. These measures are relatively not demanding of data and are easy to interpret. However, there are serious limitations to these measures. These measures imply that all opportunities are equally desirable within the limit, regardless of the time spent on travelling or the type of opportunities. The measures include elements from the land-use and transportation components, but fail to evaluate their combined effects (Geurs and van Wee 2004). Furthermore, they ignore everything outside the threshold. These measures are often used to measure local accessibility where distance is of less concern, although some success has been shown in using these measures to capture both local (Waddell and Nourzad 2002) and regional (Srour et al. 2002) accessibilities in residential location choice models.

2. Gravity-based Accessibility Measures

A typical form of measuring regional accessibility is a gravity-based accessibility measure that weights activities (e.g. number of jobs) by impedance, often as a function of travel time or travel cost. Accessibility, $Access_i$, for residents of zone *i* to jobs in all the zones in a region can then be measured as shown in equation (2) below:

$$Access_i = \sum_j Jobs_j \, f(Time_{ij}) \qquad (2)$$

Where $Jobs_j$ is the jobs (or other activities) in zone *j*, $Time_{ij}$ is the travel time (or other forms of travel cost) from zone *i* to zone *j*, and $f(Time_{ij})$ is an impedance function that can be estimated for various mode of travel as well as for various trip purposes to reflect people's level of willingness to travel. Although the impedance function may take many forms, the negative exponential form has been the most commonly used, that is, the measurement of distance separating various areas should be raised to some power.

Accessibility increases with the amount of activities increases. Accessibility also increases when activities become closer to each other.

Gravity-based accessibility measures have been successfully used by researchers to evaluate the relative ease of reaching jobs, i.e. work-place accessibility, as well as the relative ease of reaching other services such as retail stores in a metropolitan region, i.e. nonwork accessibility, and by various travel modes, including automobile and public transit (Hansen 1959; Ingram 1971; Wilson 1971).

One criticism to gravity-based accessibility is that in general it only takes "supply side" into consideration, whereas the "demand side" (e.g. how many people looking for jobs) is not considered (Morris et al. 1979). But research advancements have overcome this weakness. Shen (2001) developed a model that incorporated characteristics of job seekers such as level of education and vehicle availability. The study concluded that for job seekers who depend on public transit, very few residential locations will allow them to have an above-average access level.

3. Utility-based Accessibility Measures

Other forms of accessibility measures have been developed and used in research. One type of utility-based measure takes a form similar to that of gravity-based measures but uses the "logsum" from the mode choice model as a composite measure for impedance (Waddell and Ulfarsson 2003).

Based on random utility theory, the probability of an individual making a particular choice depends on the utility of that choice relative to the utility of all choices. When making mode choice decisions, it is assumed that an individual assigns a utility to each mode and then selects the alternative which maximizes the utility. The denominator of the mode choice model, called “logsum”, has been used to represent the impedance for calculating accessibility.

A typical mode choice model is a probabilistic model as shown below in equation (3):

$$P(i) = \exp(U_i) / \sum_j \exp(U_j) \qquad (3)$$

Where: *P(i)* = probability of choosing travel mode *i*

Ui = utility of mode *i*

The *Ui* represents the utility of a mode compared to other modes and is a function of travel costs and other attributes. Equation (4) below is the utility function of auto mode in a SEMCOG mode choice model:

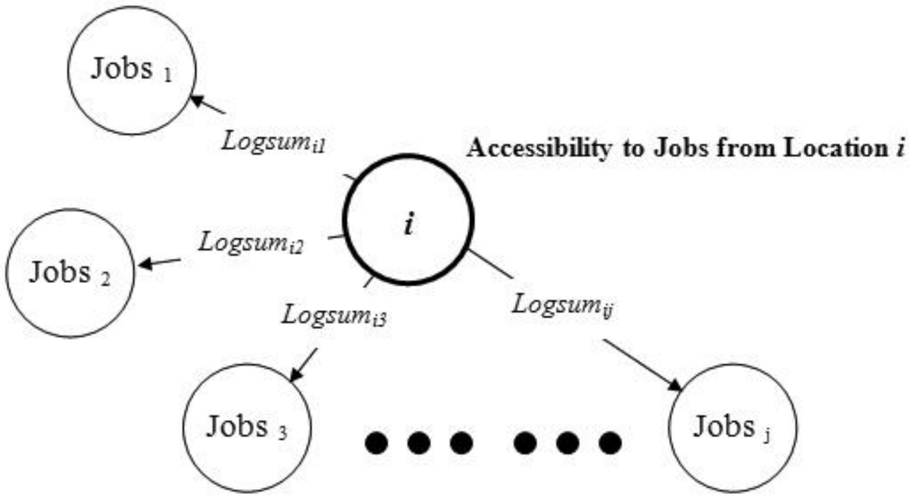
$$U_{auto} = -0.052(\text{Travel Time}) – 0.0041(\text{Auto and Parking Cost}) –5.324(\text{Worker/Auto}) \qquad (4)$$

Utility functions can be much more complex than equation (4) by taking into account of mode attributes such as transit fare and outside waiting time, as well as characteristics of travelers. Overall, it is mode specific and typically a negative number because it represents cost. The sum of utilities of all modes, a.k.a, logsum, becomes a composite measure of impedance. Equation (1) can then be transformed into the following form in equation (5).

$$Access_i = \sum_j Jobs_j e^{\text{logsum}_{ij}} \qquad (5)$$

Since logsum =< 0 (at least in theory, rescaling required if positive in practice), the exponentiated logsum is between 0 and 1, where “1” indicates best access, whereas “0” means no access. This concept is illustrated graphically in Figure 7.

Figure 7. Accessibility Measure Using Logsum



At each destination zone j, activities are typically measured the same way as gravity-based accessibility measures. For example, to measure households' accessibility to employment, number of jobs is used to represent attractiveness at destination zones, j (Waddell et al. 2008).

Just like gravity-based accessibility measures, utility-based accessibility measures are mostly used to measure regional accessibility than local accessibility. The main advantage of the logsum term is that it provides a single composite measure of travel disutility across modes, and incorporates all the factors considered in the mode choice model that have an impact on utility, such as in- or out- vehicle travel time, wait time, transit fares, tolls, comfort, etc. Its behavioral approach reflects people's preferences. But it also has disadvantages. Logsums are difficult to interpret, compared to simpler measures. Shortcomings associated with gravity-based accessibility measures may also apply to this measure.

Applications of these accessibility measures to various modes and purposes of trips will be further reviewed in the following sections. Overall, existing accessibility measures provide good indicators for neighborhoods and regions. They reflect well the impacts of land use and transportation systems on accessibility. Recent research shows that it is also possible to compare accessibility among multiple regions (Grengs et al. 2010). However, additional perspectives on accessibility continue to emerge, that provides new challenging research opportunities.

Conventional regional accessibility measures are indifferent to the clustering of regional destinations. Thus the regional accessibility impact of an isolated destination at impedance X is identical to that of a similar destination at impedance X that is part of a major regional cluster. However, it is possible that more clustered destinations have greater attractions in the region. People may be significantly unwilling to travel to a destination that is not a part of any major regional clusters. Therefore, the upward bias may exist in this situation of using conventional regional accessibility measures.

The effects of clustering on accessibility may be further explained by the "trip chaining" phenomenon. People often link two or more trips together before a return to home. This could happen to a person making multiple nonwork stops (shopping, school, services, etc.), or to a commuter making nonwork stops during a commuting trip. This "trip chaining" behavior could be driven by resource-saving in terms of shortened travel time, miles of road traveled, and the associated reductions in fuel used and other costs to fulfill a prescribed combination of activities (Southworth 1985). Therefore, selection of places to visit is likely to be affected strongly by the ability to link a given site to other sites on a multi-trip chain. The importance of such chaining in people's daily lives is supported by a large body of empirical evidence (Hanson 1980a; Horowitz 1982). Thus clustering of destinations could be essential for reducing costs of reaching those intervening destinations, hence the higher accessibility in clusters. On the other hand, destinations that are not clustered could significantly increase the cost to travel between intervening destinations in a tour. Furthermore, the importance of clustering may increase when destinations are farther away. Research has found that the likelihood of trip chaining increases as travel distance, travel cost, or the density of opportunities increases but decreases with the speed of travel (Nishii et al. 1988). Under the trip chaining logic, destinations that are along the route could be equally valuable as destinations that are in the cluster at the end of the tour, as long as a traveler is able to stop easily. One might think that routes matter a lot. But there is a significant penalty for stopping, particularly when traveling long distance at high speed. Therefore this dissertation focuses on clustering.

It is logical to think that the conventional measurement of regional accessibility needs to be tempered by an awareness of the clustering effects. A weight might be estimated for each destination based on the degree of clustering. The degree of the regional accessibility impact might depend on these weights as shown in the following equation (6):

$$Access_i = \sum_j k_j \bullet Jobs_j \bullet f(Time_{ij}) \qquad (6)$$

Where k_j is the weight of clustering for zone *j*.

To the knowledge of the author, this is the first time that clustering effects are being considered explicitly in analyzing accessibility.

D. Accessibility by Trip Purpose

While accessibility to work dominated earlier research activities, studies on accessibility for nonwork purposes have been increasing as nonwork trips increased to constitute approximately three-quarters of urban trips and also represent an increasingly large proportion of peak period trips (Rajamani et al. 2003). Work accessibility and nonwork accessibility have clear differences. The most important difference is that people are willing to travel farther to work than they are for all other nonwork purposes. Early empirical estimations of gravity models resulted in exponent values for travel time ranging from 0.5 to almost 3.0 for different trip purposes (Hansen 1959): "These studies indicate decreases in the exponent as trips become more important, i.e., school trips 2.0+, shopping trips 2.0, social trips 1.1, work trips 0.9. Inasmuch as distance appears in the denominator of the gravity model, a decrease in the exponent means that distance becomes a less restrictive factor" (Hansen 1959, p. 74). People can have shorter trips to nonwork destinations than to work places often because nonwork destinations are more ubiquitous than work places. Retail stores and service centers for daily life exist in many places at the same time.

In a much more recent study, Grengs et al. (2010) evaluated metropolitan regions in terms of the level and distribution of accessibility across the population for four accessibility categories: work accessibility by car, work accessibility by transit, nonwork accessibility by car, and nonwork accessibility by transit. A common form of the gravity model was used in this study. Distance-decay coefficients in these gravity models, representing the resistance of travel between zones, were estimated separately for work and nonwork trips. The results (0.068 for work and 0.220 for nonwork) confirmed that people are more willing to travel long distance for work than to nonwork activities.

It is important to consider both work and nonwork trip purposes in accessibility analysis. Journey to work data is frequently analyzed because it is typically more readily available and has been considered critical in travel behavior and location choice decisions. Nonwork trips are analyzed because of their increasing shares of metropolitan travel and considered representing trip types most directly influenced by levels of neighborhood access (Krizek 2005). Furthermore, work trips and nonwork trips are more connected to each other in reality than in existing research. If a commuter stops at a daycare facility and then proceeds to work or coming home, the tour that he or she is taking has home-based-nonwork, nonhome-based-work, or nonhome-based-nonwork trips in it. The primary purpose of the tour is not obviously identifiable. Therefore, any analysis that does not consider the trip-chaining phenomenon in this case would miss the unique effects and interaction of the two trip purposes. Three decades ago, Hanson (1980b) stressed the importance of jointly analyzing work and nonwork travel because separating trips by type fails to capture linked and multi-purpose travel behavior that often exists. In regard to analyzing accessibility and location choice, examining only individual trips instead of the larger pattern of linked trips in tours fails to take into account the physical and temporal relationship of intervening destinations and may provide an incomplete account of accessibility and its impacts on travel behavior and location choice.

Another issue often related to trip purpose is the selection of trip ends, as indicated in the earlier discussion. By far the most commonly used trip end in accessibility measurements is the home (Handy and Niemeier 1997). Accessibility is generally measured between a residential location and a destination. Home-based accessibility measures can be related to the demographic and socio-economic characteristics of the residents (Peng and Dueker 1995). This enables the identification of population and household characteristics with various preferences as well as identifying concentrations of vulnerable social groups. Analysis can be made in regards to how well the transportation needs of these people are being met. However, the percentage of home-based trips has been declining while percentage of nonhome-based trips as well as trip chaining has been increasing (Hu and Reuscher 2004). This calls for a growing need for nonhome-based accessibility measurements. For example, work-based accessibility measures can serve as indicators to monitor the progress in bringing the labor force closer to jobs (Cervero et al. 1999), which is a goal of jobs-housing balancing strategies (Levine 1998). In regard to equity, public transit accessibility to and from work places is especially important to transit dependent populations. Of course, it is also important to consider nonwork transit accessibility for the same population, as it may be particularly relevant for no-car households.

This dissertation aims to take into account the effects of both work and nonwork trips and their interactions by 1) including *clusters* of both work and nonwork destinations in measuring accessibility, and 2) estimating *mid-range* accessibility's effects on residential location choice. First, chained work-nonwork trips may benefit from clustered destinations, because clustering reduces travel time between intervening destinations. Furthermore, mid-range accessibility may reveal the significant unique effects of nonwork trips on location choice, as previously discussed that the biggest difference between nonwork trips and work trips is that people are less willing to travel long distance for nonwork trips. Another way to think about this phenomenon is that we need to be aware of such factors as *non-use* of remote irrelevant activities when measuring regional accessibility. For example, because there is a grocery store nearby, the existence of other grocery stores is largely irrelevant. For activities that are near, whether clustered or not, they are important for measuring accessibility. It may be argued that adjusting regional accessibility based on clustering may not be sufficient to account for all the impact of accessibility at various scales on location choice. Adding mid-range accessibility may explicitly account for the scale effects. People only use parts of the city where they live. These "parts of the city" (or parts of a region in the modern urbanized areas) could be sub-regional market areas. A sub-regional accessibility measure is therefore needed to better explain people's location choice. Accessibility may also need to be tempered for a second type of *non-use* that is based on individual's preference, wants, tastes, as well as constraints. This study deals with the first type, i.e., the geographic or market scales of the issue.

E. Place-based vs. People-based Accessibility

Most of the discussion in the preceding sections was about "place-based" accessibility measures. These measures describe level of accessibility to spatially distributed activities from various locations. However, people live in same locations may have different accessibility. An alternative to place-based accessibility is “person-based” accessibility, where accessibility is analyzed at individual level.

Person-based accessibility measures are founded in the time-space geography that was first developed by Hägerstrand (1970). Hägerstrand demonstrated the dialectical relationship between space and time by introducing the time-space prism theory. A time-space "prism" is the set of all points that can be reached by an individual given a maximum possible speed from a starting point in time-space and an ending point in time-space. These time-space prisms can be regarded as accessibility measures. They give the potential areas of opportunities that can be reached given predefined constraints.

Hägerstrand identified three interrelating sets of constraints: capability constraints, coupling constraints, and authority constraints. First, capability constraints limit the activities of an individual due to his or her biological construction or the tools he or she can command. Secondly, coupling constraints are defined as where, when, and for how long an individual has to join other individuals, tools, and materials in order to produce, consume, or transact. When an individual needs to join other individuals, tools, and materials, his or her path in space-time has to be grouped with their paths, or ‘bundle’ as Hägerstrand calls these groupings of paths. Finally, authority constraints are ‘control areas’ or ‘domains’, that are time-space entity within which things and events are under the control of a given individual or a given group.

Built upon the time-space prism theory, person-based accessibility attempts to measure accessibility from the viewpoint of individuals incorporating spatial and temporal constraints. In other worlds, these measures analyze whether and how observed or assumed individual activities can be carried out in time-space prisms under capability, coupling, and authority constraints. For example, a person who leaves work at time X at location A needs to arrive home at time Y in location B. He or she has (Y-X) amount of time to take part in other activities between and around location A and B. The more destinations he or she can reach, the higher the accessibility. If he or she can move quicker, and/or if the destinations are closer, there is a better chance for the person to enjoy higher accessibility. Accessibility measured at personal level has theoretical advantages compared to place-based accessibility measures. First, it represents the individuals' experiences on the accessibility instead of assuming that all individuals in one zone have the same level of accessibility. Secondly, it takes into account the fact that many trips that contribute to individual accessibility are made in the context of the sequential unfolding of an individual's daily activities, i.e., trip chaining may be considered explicitly at the personal level in the measurements. Finally, it considers time-space constraints that may render many opportunities in the urban environment unreachable by an individual (Makri and Folkesson 1999).

Time-space theory seems very promising in measuring personal accessibility because of its disaggregated approach. However, their applications in accessibility research have been relatively rare. Using Hagerstrand's time-space prism theory, Lenntorp (1976) developed maps to determine the parts of an individual's environment that are physically accessible, or "within his physical reach" in Lenntorp's words. Kwan (1998) developed a distance matrix for all locations in her study area using the shortest path algorithm. The feasibility of each network link is then tested by going through the entire matrix and identifying those links that are reachable within the time-space constraints for any given pair of fixed activity locations. The results demonstrated that time-space based measures capture activity-based contextual effects which are not incorporated in traditional location-based accessibility measures. This allows more sensitive assessment of individual variations in accessibility, including gender and ethnic differences. Lee et al. (2010) developed an accessibility measure to nonwork destinations based on the time-space prism theory using data from the central Puget Sound region. They used a traffic analysis zone (TAZ) system. First, they assessed daily activity schedule of an individual using household travel survey data. Then they determined for each TAZ whether it could be visited with sufficient time given a travel mode. Finally they applied the cumulative opportunity approach to the set of TAZs to quantify the opportunities that are available to an individual given the temporal and spatial constraints. Their results show that person-based accessibility to nonwork destinations is statistically significant in residential location choice.

Although person-based accessibility can reveal individuals' specific experiences in their socio-spatial context by measuring accessibility enjoyed by a particular person having specific needs and resources, it also has disadvantages. The biggest disadvantage is the difficulty of operationalizing the concept. This is troublesome in several ways. First, space-time measures require large amount of information about detailed individual travel and activity data which are not typically available. Second, despite advancements in programming, GIS, and spatial modeling, operationalization of person-based accessibility measures still faces difficulties for lack of feasible operational algorithms. Third, even with today's computing power, the computational intensity still makes it difficult to use space-time measures in large-scale projects. The applications are often restricted to a relatively small area or corridor and to a subset of the population because of the large data requirements and computing intensity. Finally, the results are difficult to aggregate to evaluate accessibility to a larger population and to a bigger geographical area (Geurs and van Wee 2004).

Although this dissertation does not deal with the full range of person-based accessibility, it considers one simple type but important aspect of personal accessibility, that is individual workers' commute time from home to work.

F. Accessibility by Mode

While much of existing research on accessibility dealt with automobile accessibility, a considerable number of studies focused on public transit or non-motorized (e.g., walking and biking) accessibility.

1. Transit Accessibility

Developing accessibility measures for transit could be especially complicated because of its inherent spatial and temporal characteristics. Existing public transit accessibility measurements fall into two broad categories: 1) accessibility that measures access to a public transit system, and 2) accessibility that measures public transit access to destinations.

a) Access to public transit system

When measuring public transit accessibility, many measurements actually are only concerned with the ease of accessing transit stops and stations. In other words, these measurements are assessing the catchment potentials of transit stops and stations. These measurements usually do not assess the destinations and activities that the transit users can actually reach. The underlying assumption is that access to the public transit system acts as a proxy for access to a range of destinations (Kerrigan and Bull 1992). These assumptions, of course, are only valid when the public transit system is really well designed and implemented.

The simplest measurement of this type is the area buffer method that is used to define the extent to which a public transit system can reach its users. Area catchment maps can be created by drawing circles around transit stops and stations. The radius that has been used most often is around one quarter mile or 400 meters, which has been found to be the distance that people are willing to walk to use public transit service, a.k.a., “walking distance” (Murray 2001). Once the area that public transit system can serve is determined, the proportions of population, households, workers, jobs, and other activities within walking distance to transit service can be calculated. This simple area buffer method can be easily implemented with GIS. The results are also easy to interpret. But it has obvious shortcomings.

First, using a Euclidean distance to create the buffer could overestimate accessibility, because the Euclidean distance represents the farthest reach of actual walking distance from a stop or station. The difference between this farthest reach and actual reach may vary significantly depending upon the configuration and connectivity of the pedestrian network leading to a stop or station. These buffers ignore physical barriers to walking and tend to overestimate the spatial service coverage of a transit system (Horner and Murray 2004). Secondly, using area ratios to calculate the number of people or jobs that can be served by the transit system may have bias too (Zhao et al. 2003).

Area buffers often overlap with census tracts, transportation analysis zones (TAZs), zip codes, or any other areas where population and employment data are based on. When calculating severed population or jobs using area ratios, even distribution is assumed within these areas although they are often too large to ignore variations within them. The measurement would be more accurate with more disaggregated data such as land parcel data.

To overcome the shortcomings of area buffer method and area ratio allocation, the network ratio method was proposed by utilizing the layout of street network (O'Neill et al. 1992). In this method, population and jobs are allocated based on the proportions of street length within a buffer. The underlying assumption has changed from even distribution within a buffer to that number of residents and jobs on a street are proportional to its length. Even though this is still a weak assumption, but it is no weaker than the assumptions in the area ratio method. Other studies using the network ratio method confirmed that this method is better suited for measuring access to a transit system than area ratio method (Hsiao et al. 1997).

Attempting to tackle the weakness of both area ratio and network ratio methods, some researchers developed other methods that mostly eliminated the allocation process. Zhao et al. (2003) used disaggregated parcel data to account for population and job distribution. They also explicitly considered physical barriers such as water bodies and walls in the pedestrian network. After comparing the results from this parcel level measurement to those from area and network ratio methods, they found the previous methods constantly overestimated catchment potentials of public transit systems.

While measuring the physical dimension of access to transit progressed from area ratio, to network ratio, to parcel-level methods, the temporal dimension of access to transit was also added to accessibility measurements in other studies. The temporal perspective is about how often transit users are served by the system. The goal is to not only identify areas that are within walking distance to a transit stop or station, but also estimate the level of service in terms of average waiting time or frequency of service in order to measure accessibility to transit more accurately (Polzin et al. 2002).

b) Accessibility to Destinations by Transit

A different set of transit accessibility measurements actually measures how well destinations can be reached by public transit systems instead of measuring access to the system only. Most of these measurements use the gravity-based method that weights the quantity of opportunities at the destination zones by impedance as a function of travel time or cost from the origin zones. Based on these gravity-based models, the more opportunities there are in the destination zones and the shorter time spent on public transit to get there from the origin zone, the higher the accessibility. Two recent examples of this type are from Grengs (2010) and Lee (2009).

Grengs (2010) used a modified gravity model (Shen 1998) to measure accessibility of low-skill workers to low-wage jobs in Detroit region. This modified gravity model has two advantages. First, it considers not only the supply of jobs but also recognizes that the workers who compete for jobs are not evenly distributed in space. In other words, it accounts for the spatial difference in job demand. Secondly, it measures accessibility by mode, i.e., measuring accessibility by auto and accessibility by transit side by side. Accessibility of residents in each zone traveling by automobile and by transit is calculated explicitly. The application of gravity-based accessibility measurement in his study clearly demonstrated the similarity of measuring accessibility by transit and by automobile.

Lee (2009) developed a prototype measurement to estimate public transit accessibility at the parcel-level that takes into account some of the most advanced development in both measuring access to transit system and access to destinations by transit. This method uses a land use classification system at the parcel-level to measure accessibility, in terms of total transit travel time, to each location of a destination type from land parcels that are served by a transit system. The accessibility measure can be of a single type of destination (e.g. retail), or it can be combined with the accessibility measurements of other types of destinations to produce a composite measure. On the origin side (e.g. residential), it may be categorized by the same land use classification system to specify certain types of trips (e.g., home-based or work-based trips). Ultimately, this technique can be used to measure the accessibility between any two types of land uses. In regard to travel time, it takes into account of in- and out-of-vehicle transit travel times to compute the accessibility measure. This includes travel times in transit vehicles; walk access from the origins to the transit network, as well as from the transit network to the destinations; transfer walk times if the transfers do not occur at the same stops; and average waiting times for transit vehicles to arrive at the stops or stations. The prototype measurement that calculated transit access to grocery stores in Seattle region showed promising results, although it required intensive computing.

Summarizing the above two types of transit accessibility and using an analogy to automobile accessibility, one can think that measuring access to transit system is equivalent to measuring local accessibility using cumulative opportunities method with transit stops and stations specifically as destinations, whereas measuring transit access to destinations is similar to measuring regional accessibility in automobile mode by using gravity-based measurement. Clustering of destinations would still matter, although it is not so much about have closer transit stops, but rather it is mostly about clustering destinations around stops, that makes chained trips, e.g., transit to walking, more efficient. Sub-regional accessibility may also be significant in transit mode, for passengers are less willing to travel long distance on a bus or train for nonwork purposes.

2. Non-motorized Accessibility

When measuring non-motorized accessibility, measurements need to be sensitive to detailed local conditions. Small or disaggregated zonal systems are required to reveal variations in accessibility. Accessibility by non-motorized mode is mostly described by local, or neighborhood, accessibility measures. Existing research showed that accessibility increases when 1) increase local concentrations of population and employment, 2) encourage a mix of appropriate land uses, and 3) design development and street network improvements to be pedestrian oriented. These factors can be summarized into three words: density, diversity, and design, or "3Ds" (Cervero and Kockelman 1997). Some additional measures attempt to combine two or all three of these categories to develop a “composite” accessibility index.

a) Density

Neighborhood density is the most readily available urban form measurement to operationalize non-motorized accessibility. It is more commonly used than any other local accessibility measures (Steiner 1994). However, it needs to be used with caution. A common practice of density calculation is to separate residential density from non-residential, e.g. commercial, industrial, and institutional densities. But it is really the combination between the two that affects accessibility most. A large neighborhood of high-density residential development only helps little to promote walking or biking for lack of attractive destinations in the neighborhood. Land use mix is critical.

b) Land Use Mix

A mix of land use can promote walking or biking as a substitute for auto travel. But it requires the right mix, meaning that land uses complement one another in terms of functions. Banerjee and Baer (1984) identified that the most valuable land uses that people value in close proximity to their home are drug stores, food markets, gas stations, post offices, specialty food stores, and banks. Furthermore land uses that complement one another need be close enough to encourage walking, biking, and other non-motorized trips. Measures to capture non-motorized accessibility range from simple inspection of presence or absence of nonresidential uses in a residential neighborhood, to counting employment in the neighbor, to complex index of land use dissimilarity in the neighborhood (Cervero and Kockelman 1997).

c) Design and Street Pattern

While density and mixed uses are often used as local accessibility measures, there are other physical characteristics that affect non-motorized accessibility. Lynch (1962) identified design features that have shown to be critical to quality experience of pedestrians, cyclists, and transit users in the built environments, including sidewalks, building scale, streetscape, and landscaping. It is obvious that operationalizing all these concepts are difficult. Some studies used gridded streets as a surrogate to measure traditional neighborhood characteristics (Cervero and Gorham 1995). Traditional neighborhoods with gridded streets contain characteristics that make walking, biking, and transit more attractive compared to modern subdivisions with cul-de-sac type of roads. McNally and Kulkarni (1997) performed a cluster analysis and found that gridded streets were one of the most influential variables to separate traditional and suburban neighborhoods.

d) Composite Index

Density, diversity, and design may re-enforce each other or undermine each other's effects on accessibility in neighborhoods. A neighborhood with density and sidewalks but no diversity in land use may or may not provide good walking accessibility. Aimed to take into account multiple dimensions of local accessibility, some researchers attempt to develop composite index. Krizek (2003) developed an index that considered all three dimensions. Using Puget Sound Regional Council (PSRC) data, for each 150-meter grid cell in the Seattle region, Krizek calculated housing density, number of employees in neighborhood retail services, and number of street intersections. Then a factor analysis reduced these three variables into a single index.

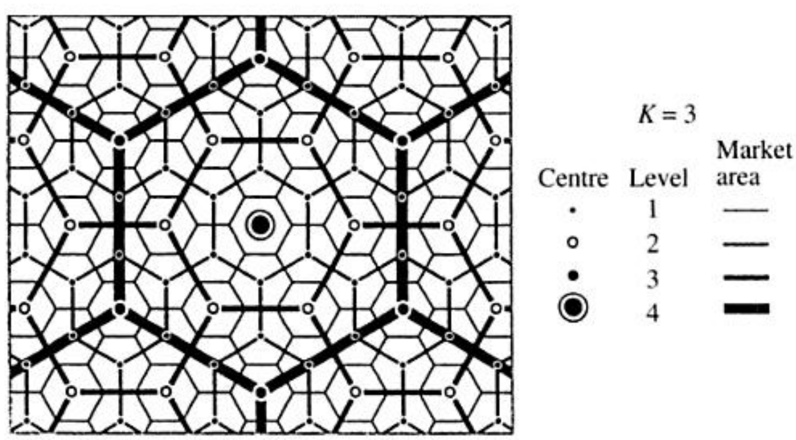
Local accessibility was originally used for exploring land use's impacts on travel behavior (Crane 2000). For example, higher density development may reduce the percentage of trips taken by auto. Increasingly, local accessibility is also used as an independent variable in urban modeling applications to predict location choice of households (Waddell and Nourzad 2002).

G. Central Place Theory

Finally, the hierarchy of market areas in a region may help further explain why local, regional, and “mid-range” accessibility may all have their own and independent impact on residential location choice.

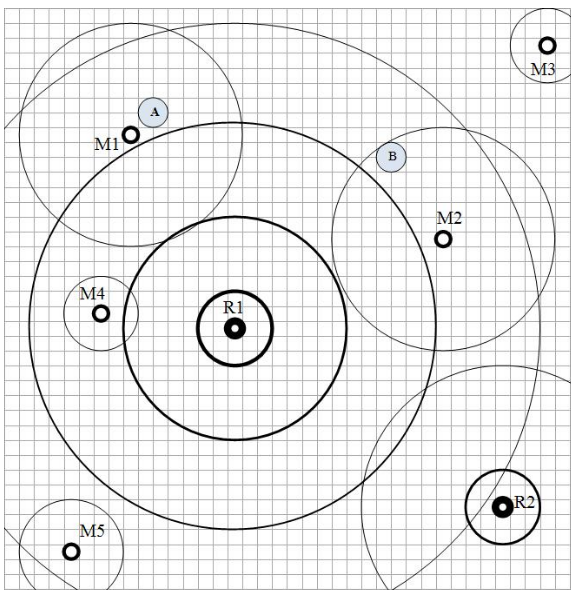
Seeking to explain the number, size, and location of human settlements in an urban system, Christaller (1933/1966) constructed a hierarchical central place theory, as illustrated in Figure 8. It asserted that settlements functioned as 'central places' providing services to surrounding areas. A region is consisting of nested market areas of various economic scales. The higher the order of the goods and services, the larger the range of the services, and the longer the distance people are willing to travel to acquire them. Hence the residents trade transportation cost for higher level of services. Examples for high order goods and services are: jewelry stores, shopping centers, and sports venues. They are supported by a much larger threshold population and demand. Examples for low order goods and services are: groceries, barbershops, and fitness centers. They are supported by a smaller threshold population and demand.

Figure 8. Central Place Theory: A Case of Christaller Model



Now, consider two households A and B in a region as shown in Figure 9, where A is located near the center of sub-regional market area M1, while B is located further away from the center of sub-regional market M2.

Figure 9. Accessibility at Multiple Scales in a Hierarchy of Market Areas



Using conventional regional accessibility methods such as gravity-based measures, regional accessibility of A or B takes into account all activities in the region dampened by travel impedance. Activities at all regional centers R1 and R2 as well as all sub-regional market areas M1 to M5 are included. However, the hierarchy of goods and service provision implies multiple levels of accessibility.

At the regional level, places that provide the highest order of goods and services, i.e., R1 and R2 in this chart, are more desirable destinations where A and B are willing to travel for long distance.

For the lower order of goods and services, accessibility matters most at the sub-regional market level. If there is a grocery store in each market area, what matters to household A is the accessibility to the store in M1. All other grocery stores are mostly irrelevant. Including all activities in the entire region without considering the order of services as in the conventional regional accessibility measurement could over-estimate accessibility in general. Furthermore, assuming household A has greater accessibility than B at the sub-regional market level (This could be because of denser activities in M1, closer destinations in M1, higher impedance in M2, longer distance of B to sub-regional center, or combinations of these factors.), if conventional regional accessibility were used for A and B indifferently, it could disproportionally over-estimate accessibility for B.

In addition, previous studies have shown that CBDs and places close to CBDs have much higher accessibility for work, shopping, social, and recreational trips (Shen 2001). While this is generally expected, it might be an indication that sub-regional accessibility has been under-represented in conventional accessibility measures for today's mostly decentralized regions with substantial secondary market centers.

While regional and local accessibility are important components of the hierarchy of accessibility, a mid-range sub-regional accessibility might be able to make the hierarchy more complete in accessibility measurements. At the neighborhood scale, as indicated by the lightly shaded circles in Figure 9, local accessibility can be used most effectively for measuring the effects of non-motorized modes as opposed to motorized regional accessibility. Effects of the sub-regional areas surrounding the neighborhoods might be measured by a mid-range accessibility that could be most effectively used for accounting the effects of nonwork trips which are particularly significant for household daily needs and are more sensitive to travel distance than work trips (Figure 2).

The research question that this dissertation attempts to address is as follows. First, can regional accessibility measurements take into account the effect of clustering of regional attractions to be more accurate indices for residential location choices? Secondly, does a mid-range sub-regional accessibility have its own significant effect on residential location choice when controlling for local and regional accessibility? If significant positive coefficients are found for sub-regional accessibility in residential location choice model, it would indicate that households are more likely to locate in sub-regional market areas that provide higher accessibility. Thirdly, how do place-based accessibility measures and individual commute time affect residential location choice?

H. Summary of Chapter II

Accessibility is a multi-dimensional concept. Dimensions of accessibility include: regional vs. local accessibility, place-based vs. person-based accessibility, accessibility by mode, and accessibility by trip purpose. Numerous accessibility measurements have been developed either to tailor to a specific dimension or to capture effects of as many dimensions as possible. Despite research advancements in measuring accessibility, limitations of existing measurements require continued creative thinking and innovation of new methods. Clustering of destinations may be important to account for the increasing trip chaining phenomenon, whereas a mid-range accessibility may further explain residential location choice given the uniqueness of nonwork travel. Central Place

Theory helps explain the needs of these innovations. The focal point of this study is how various accessibility measures affect residential location choice.

Next chapter will describe the data needs and provide descriptive analyses of the study area for this dissertation.

CHAPTER III

DATA SOURCES AND PROCESSING

The first two chapters of this discussion suggested that a sub-regional mid-range accessibility is needed in assessing accessibility and estimating the effects of accessibility on residential location choice. It also suggested that models incorporating clustering of activities into measures of accessibility would show stronger explanatory power in predicting residential location choice than models that do not consider clustering of activities. Furthermore, both place-based accessibility measures and commute times could affect residential location choice. The following analyses will examine the hypotheses using data from the seven-county Detroit region in Southeast Michigan.

This chapter presents an overview of the Detroit region in terms of demographic, socio-economic, land use, and transportation trends and patterns to serve as a background for analyses. The discussions on these trends are embedded in the following sections on data sources and data development that are needed for measuring accessibility and modeling residential location choices in the region. First, an overview of the region at the macro level is presented. This will then be followed by discussions on data that are developed at very detailed micro levels down to disaggregated individual households and land parcels.

Four major data sources are discussed in this chapter: (1) 2004-2005 Michigan Department of Transportation (MDOT) and SEMCOG household travel survey; (2) land parcel map and the associated property assessment data; (3) synthesized individual households and population for the entire region; and (4) employment data. The household travel survey was conducted by MORPACE International. Land parcel and assessment data, synthesizing households and population, and employment data were originally collected and processed by SEMCOG's Data Center led by this author for developing and running SEMCOG's UrbanSim forecast model. Further adjustments were made for this dissertation research.

Data development methods for UrbanSim can be found at the UrbanSim website (www.urbansim.org). Additional data procedures developed by SEMCOG staff are documented on SEMCOG's internal wiki site, which can be obtained upon request. The highly disaggregated data set used for this study allows for detailed behavior-based analysis on household's residential location choice with respect to accessibility at various levels from land parcels to the entire region.

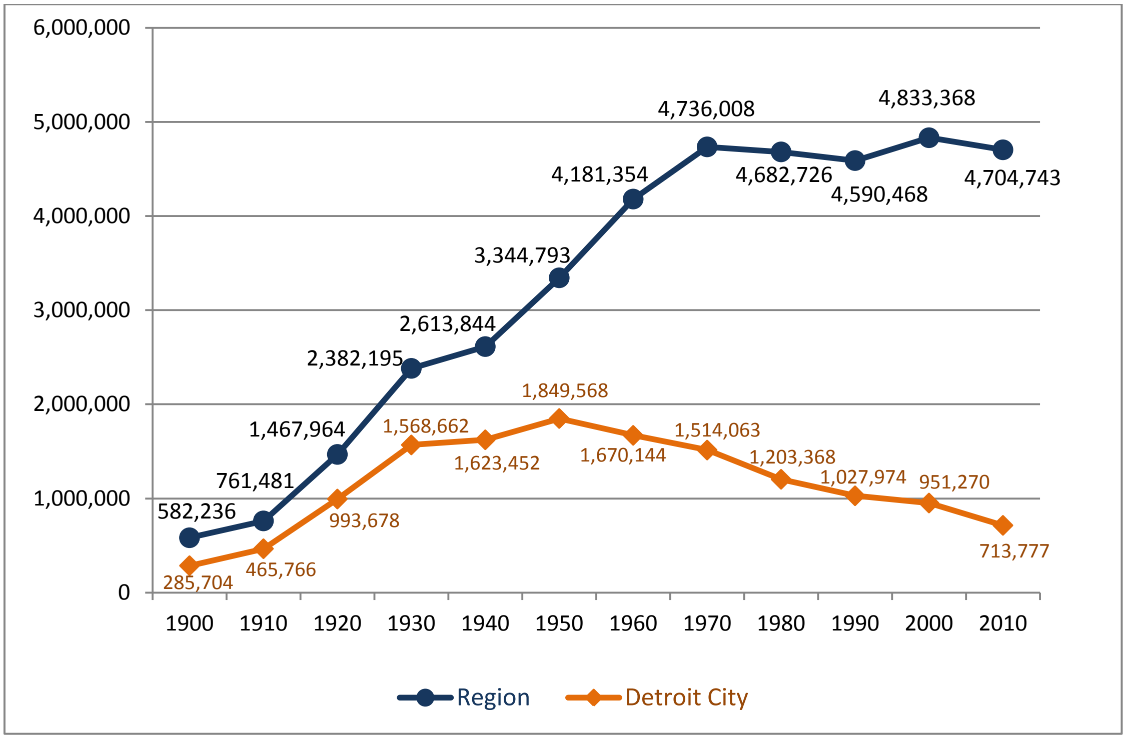
A. Overview of Study Area

The study area for this dissertation is the Detroit region, that is defined as the seven county area in Southeast Michigan where SEMCOG is responsible for regional transportation planning. Approximately 50% of population and jobs of the State of Michigan are located in this region (Figure 10).

Figure 10. Study Area - Seven Counties in Southeast Michigan

The Detroit region has experienced economic turbulence and has been in a critical transition in its recent history. Population changes tell the story of the rise and fall of the region and its central city, Detroit (Figure 11).

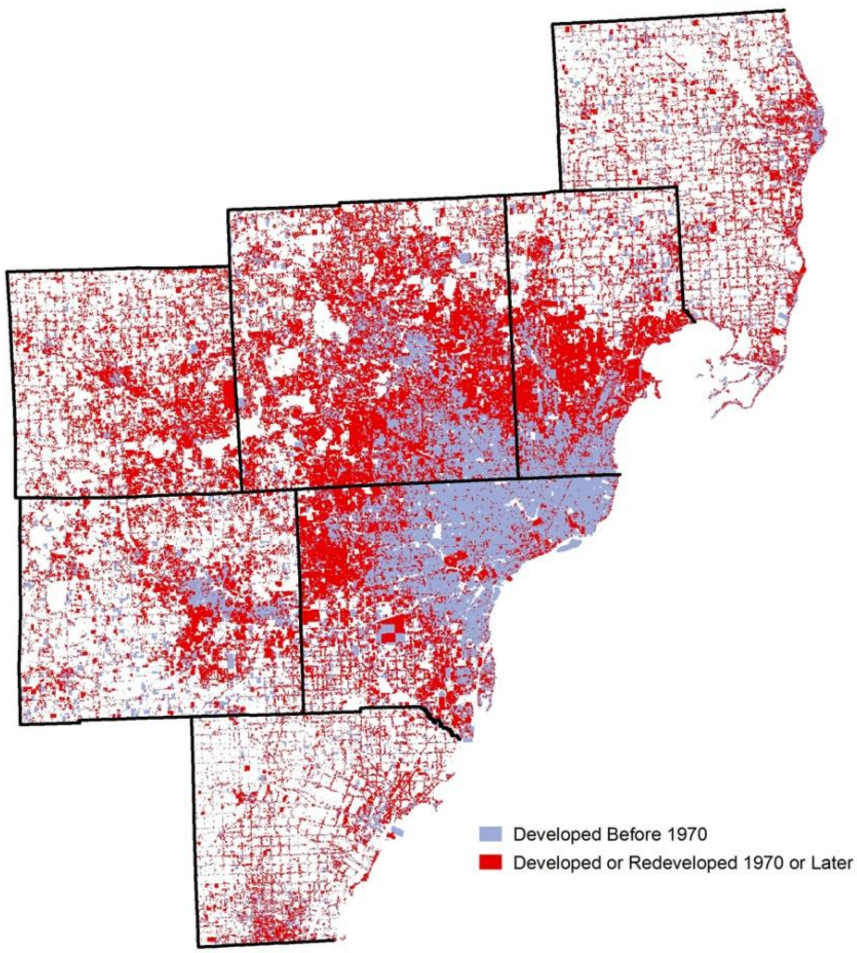
Figure 11. Population History – Detroit City vs. the Region



Data Source: United States Census Bureau

The region's population experienced little change in the last 40 years since 1970. However, distributions of population in the region, and consequently land use patterns, changed significantly in the same time period. The central city Detroit's population has been declining since 1950, down from over 1.8 million to only 714 thousand. The city's share of the region's population has declined from 32% to 15% in the same time period. Employment experienced an even larger shift away from the central city. Only 13% of the region's jobs were in the City of Detroit by 2005 (SEMCOG 2005). Because population and employment continued to shift to suburbs, significant amount of land has been developed outside the central city, while a large part of once developed land has become vacant in the City. The developed land in the region doubled from 1970 to 2010 while population remained virtually unchanged (Figure 12).

Figure 12. Land Developed before and after 1970



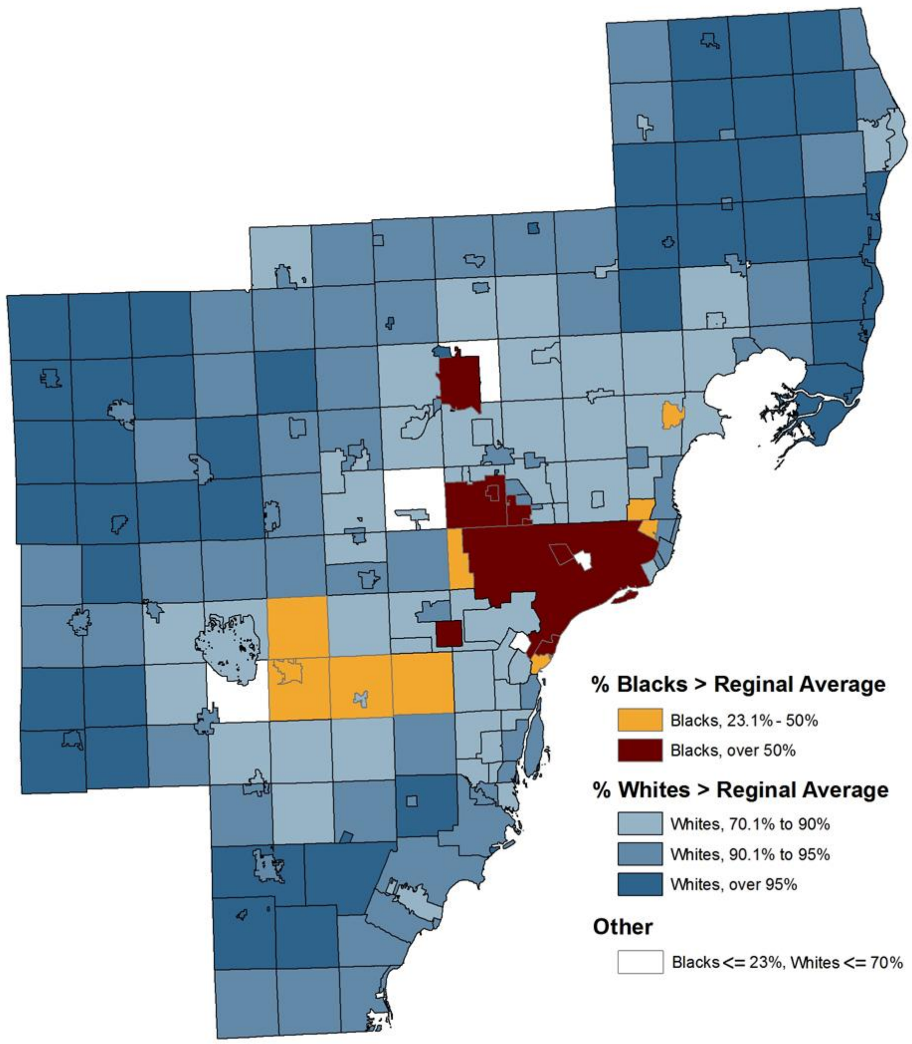
The Detroit region has long been transformed from a Detroit-dominated monocentric metropolitan area to a much more sprawling region with numerous sub-regional centers. Residential location choices reflect the preferences of households of various demographic and socio-economic characteristics as well as the changing physical form of a region. People's travel behavior has also changed, including the frequency of trip-making, the length of the trips, and the mode of travel for both work and nonwork purposes. Some of the key household characteristics and travel patterns are further discussed in the following sections when data sources are discussed in greater detail. But racial patterns of residential location are discussed first here because they are such dominant shapers of accessibility in the region.

The Detroit region is one of the most racially segregated areas in the United Stated. Frey and Myers (2005) ranked Detroit metropolitan area the second highest segregated region using a dissimilarity index for Blacks and Whites (Frey and Myers 2005, p. 39, Table A2). The concentration of Blacks and Whites as well as the separation of the two races are shown in the following map (Figure 13).

The region has 4.7 million people (Census 2010). The White population accounts for nearly 70 percent of the total population, whereas the Black population makes up 22 percent of the total. While the total Black population exceeds 1 million, most live in a few concentrated municipalities. There are 235 municipalities in the region. Approximately 80 percent of Black population lives in the 19 municipalities where percent of black population is higher than the regional average. Two-thirds of Black population lives in just five municipalities. The City of Detroit is home to over 586,000 or 58 percent of total black population in the region. The Black population makes up 82 percent of total population in the City of Detroit.

Meanwhile, there are 211 municipalities where the percentage of the White population is higher than the regional average of 70 percent. There are 147 municipalities where the White population is over 90 percent, and there are 73 of them where the White population is over 95 percent. The City of Detroit has only 55,600 Whites, which account for only eight percent of the City's total population. What this means for residential location choice modeling is that it is not enough to use variables of accessibility and affordability, etc. Race needs to be explicitly represented as independent variables, interacting with other variables. Estimating separated models for various races is also desirable.

Figure 13. Population by Race by Municipality, Detroit Region, 2010



The goal of data development for this study is to estimate household location choice models, including measuring accessibility of various forms as independent variables in the choice models. To estimate choice models, data are needed to characterize the “choosers” and the “choice set.” In this study, the “choosers” are households making location choice decisions. Those households with their demographic and socio-economic characteristics were obtained mainly from the 2004-2005 Michigan Department of Transportation (MDOT) and SEMCOG household travel survey. The “choice set” in this study is the residential buildings on individual land parcels in the Detroit region. The main source of data for characterizing this choice set is the land parcel map with property assessment data. In addition, a complete universe of households and population in the region were created to help characterize the demographic and social environment of the choice set. Employment data by Traffic Analysis Zones (TAZs) were developed for measuring accessibility to jobs. Additional input data files include U.S. Census data, regional environmental GIS files such as wetland, preserved farmland and recreation land, water and sewer service areas, and community master plans that are collected and processed by SEMCOG.

B. Household Travel Survey

As in most metropolitan areas in the United States, household travel survey is periodically conducted in the Detroit region. During 2004-2005, Michigan Department of Transportation (MDOT) conducted a comprehensive data collection program known as

Michigan Travel Counts, which collected activities and travel inventories from all members of 14,996 randomly selected households within the state of Michigan, of which 2,222 households completed were within the Southeast Michigan region. To supplement this Southeast Michigan sample, SEMCOG commissioned a separate SEMCOG Travel Counts program, as an add-on component to Michigan Travel Counts. The objective of this effort was to collect activities and travel inventories from an additional 3,843 randomly selected households within Southeast Michigan. When the MDOT and SEMCOG samples are combined, data precision for the region is improved. The combined MDOT and SEMCOG files are used for this study to analyze residential location preferences with respect to neighborhood characteristics and accessibility in Southeast Michigan.

1. Survey Design

The MDOT household travel survey was designed to collect 48 hours of travel data from approximately 2,040 households in each of seven regions in the State of Michigan. Travel diaries were sent to each member of a household for completion, regardless of age, and to any overnight visitors the household may have during the assigned travel period. In addition to collecting travel information for the assigned period, respondents were also asked to provide demographic and socio-economic data. Randomly selected households were sent a pre-notification letter informing them that they would be receiving a telephone call within the following week. Once a household agreed to participate during the recruitment call, they were assigned a two-day travel period. Travel diaries were mailed and each household received a reminder call the day before the assigned travel days. Households were then called for retrieval of their travel information the day following their travel period. If a household did not respond via telephone, survey questionnaire including the travel diaries could be mailed in or data could be entered via the internet.

The goal of the survey was to gather enough data to account for the variance in travel patterns across the entire State of Michigan to support travel demand forecast model development for the State. Southeast Michigan was one of the seven geographic regions statewide for the purpose of this survey. Because the survey sample design called for equal number of survey households located in each one of the seven regions, only one seventh of the sample households were from Southeast Michigan region. However, nearly 50% of the State's population and employment are located in this region. Furthermore, travel demand forecast models at metropolitan level typically have higher data requirements than statewide models. SEMCOG decided to add approximately 3,800 more sample households to supplement the MDOT survey that had 2,222 sample households in Southeast Michigan, using the same consultants and similar methodology, for developing a new regional travel demand forecast model.

In the supplement SEMCOG survey, additional geographic stratifications were added to ensure representative samples for smaller areas and low-response-rate areas. The sample design for SEMCOG survey divided the Detroit region into eight geographic sample areas. Each sample area is defined by a county. Wayne County is split into City of Detroit and the balance of the county. The target sample size is proportional to each area's share of total households in the region. When the MDOT survey and SEMCOG supplement survey were combined, the resultant sample had a 5% or less estimated standard error at 95% confidence level for each of the eight areas in the region. The overall standard error for the region was 1.26% at 95% confidence level (Table 2).

Table 2. MDOT and SEMCOG Survey Samples for Detroit Region

Sampling Areas	Sample Sources			Standard Error (at 95% Confidence Level)
	MDOT	SEMCOG	Total	
Livingston County	82	358	440	4.67%
Macomb County	354	612	966	3.15%
Monroe County	69	279	348	5.25%
Oakland County	562	666	1,228	2.80%
St. Clair County	103	408	511	4.33%
Washtenaw County	213	400	613	3.95%
Balance of Wayne County	421	640	1,061	3.01%
City of Detroit	418	480	898	3.27%
Total Region	**2,222**	**3,843**	**6,065**	**1.26%**

Another difference between the SEMCOG survey and MDOT survey is that SEMCOG survey has a one day travel diary instead of a two day diary in the MDOT survey. Analysis on MDOT survey results showed that there were concerns with the second day of trip dairy in the survey. For example, personal trip-rates dropped from 3.64 in the first day dairy to 3.19 in the second day dairy. Similarly, zero-trip households increased from 8.1% in the first day dairy to 11.0% in the second day dairy. Therefore, for this study, only the first day MDOT survey was combined with the SEMCOG one-day survey. A lesson learned here is that it is essential to note the differences in survey results between days in a multi-day trip dairy. One day of travel dairy is possibly sufficient for analysis needs.

The design of survey instruments, the trip diary in particular, was aimed for improving existing traditional four-step travel demand forecast models as well as for developing the next generation of activity based models. To achieve that goal, a location-based travel diary format was used. In this type of travel diary, respondents would be carefully taken chronologically through their travel days from location to location, recording both their activities at locations and their detailed travel information between locations. Analysis of trip chaining is possible based on this survey design.

2. Selected Survey Results

There are six major variables on household characteristics that can be developed for the survey results. They are listed in Table 3 with the ranges of their values.

Table 3. Household Variables from Travel Survey

Household Attributes	Values
Number of Persons (Household Size)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more
Age of Household Head	Single year, 18 to 97
Household Income	$0 to $125,000 or more
Number of Children	0, 1, 2, 3, 4, 5, 6, 7, 8 or more
Number of Workers	0, 1, 2, 3, 4, 5 or more
Number of Vehicles Available	0, 1, 2, 3, 3, 4, 5, 6, 7, 8, 9, 10 or more

Race is often a significant variable in determining residential location choices. However, race was not asked in either MDOT or SEMCOG survey. A synthesized race attribute was assigned to each household. The methodology for synthesizing households and population is explained in Section D: Synthesizing Households later in this chapter. It is similar to the method developed at the Los Alamos National Lab for the TRANSIM software (U.S. Department of Transportation 2005), but enhanced by weighting more on marginal distributions at census block-group level, and by considering local housing characteristics developed from assessment data when assigning synthesized households to parcels.

The original sample data from the surveys were expanded to reflect the entire population of the region. Data expansion factors were developed from 2000 census Public Use Microdata Samples (PUMS). The total expanded household and population numbers were compared to 2005 SEMCOG population and household estimates for validating the survey results. All the analyses presented in the remainder of this section are based on the data expanded from the survey samples to the population of the areas surveyed (Table 4).

a) Households

Table 4 provides a breakdown of various household characteristics. Household size various from the smallest of 2.41 in Washtenaw County to the largest of 2.73 in Livingston County. The more developed counties such as Macomb County, Oakland County, and Wayne County excluding City of Detroit all had smaller than regional average household size, whereas the more rural counties including Monroe County and St. Clair County had larger than average household size. Detroit led average number of children per household, followed by Livingston County where large families with more children tend to live. Washtenaw County had the lowest average age largely due to its college population. It was followed by City of Detroit where households had more than average number of children.

Table 4. Selected Household Characteristics Based on Expanded Samples

	Detroit	Living-ston	Macomb	Monroe	Oakland	St. Clair	Wash-tenaw	Balance Wayne	Region
Households (HH)	322,393	64,611	327,861	60,068	497,466	66,538	138,846	448,316	1,926,099
Percent of Total	16.7%	3.4%	17.0%	3.1%	25.8%	3.5%	7.2%	23.3%	100.0%
Persons	858,231	176,339	819,659	158,827	1,243,874	171,487	334,728	1,127,986	4,891,131
Persons/HH	2.66	2.73	2.50	2.64	2.50	2.58	2.41	2.52	2.54
Workers	298,157	90,753	402,495	76,073	626,527	82,670	179,057	541,852	2,297,584
Workers/HH	0.92	1.40	1.23	1.27	1.26	1.24	1.29	1.21	1.19
Vehicles Available	412,084	147,895	621,424	130,259	962,585	126,241	257,282	861,717	3,519,487
Vehicles/HH	1.28	2.29	1.90	2.17	1.93	1.90	1.85	1.92	1.83
Children	268,819	49,974	205,312	41,231	329,837	44,026	89,642	283,482	1,312,323
Children/HH	0.83	0.77	0.63	0.69	0.66	0.66	0.65	0.63	0.68
Average Age	36.6	37.9	38.8	38.5	38.3	38.7	36.4	39.2	38.2
Licensed Drivers	477,108	125,945	592,508	114,527	888,458	122,820	241,156	810,748	3,373,270
Licensed Drivers/HH	1.48	1.95	1.81	1.91	1.79	1.85	1.74	1.81	1.75
Driver Trips	1,461,472	461,835	2,098,148	405,933	3,423,399	433,845	888,128	2,937,913	12,110,673
Driver Trips/HH	4.53	7.15	6.40	6.76	6.88	6.52	6.40	6.55	6.29
Transit Trips	161,278	2,527	15,853	857	15,940	7,148	36,023	45,674	285,300
Transit Trips/HH	0.50	0.04	0.05	0.01	0.03	0.11	0.26	0.10	0.15
Mean Travel Time	21.5	22.3	18.4	18.7	18.5	18.0	18.1	17.3	18.8
Mean Travel Time to Work	27.9	33.4	26.1	23.5	26.7	24.7	24.2	25.1	26.2

b) Workers

Respondents were asked to provide both home address and work address. All findings discussed here are based on respondents' home address. The average number of workers per household for the region was 1.19. Only the City of Detroit had the lower than average number of workers at 0.92 per household. The number went up to as high as 1.40 in Livingston County. These workers' commute time is used explicitly as independent variables in residential location choice modeling for this study.

c) Vehicle Availability and Licensed Drivers

Households were asked the number of cars, minivans, and light trucks available to household members for travel. The number of vehicles available per households ranged from 1.28 in Detroit to 2.29 in Livingston County, which is a difference of one full vehicle per household. The number of licensed drivers is highly related to number of vehicles available. But the difference between the lowest number (1.48 per household in Detroit) and the highest number (1.95 in Livingston County) is not as large as the difference in vehicles available. This could be a reflection on the unmet needs in vehicles in Detroit and the "excessiveness" of vehicles available in Livingston County.

d) Income

Household income distribution for all eight areas is presented in Table 5. The shaded cell in each column is the income category that contains the median value for that place. Livingston, Oakland, and Washtenaw counties saw more households in the higher income categories, with approximately 50% of their households reporting income of more than $60,000. On the other hand, City of Detroit's income is significantly lower than the other areas. More than half of Detroit's households had annual income less than $40,000. St. Clair County also had larger shares of its households in the lower income categories.

Table 5. Household Income Distributions

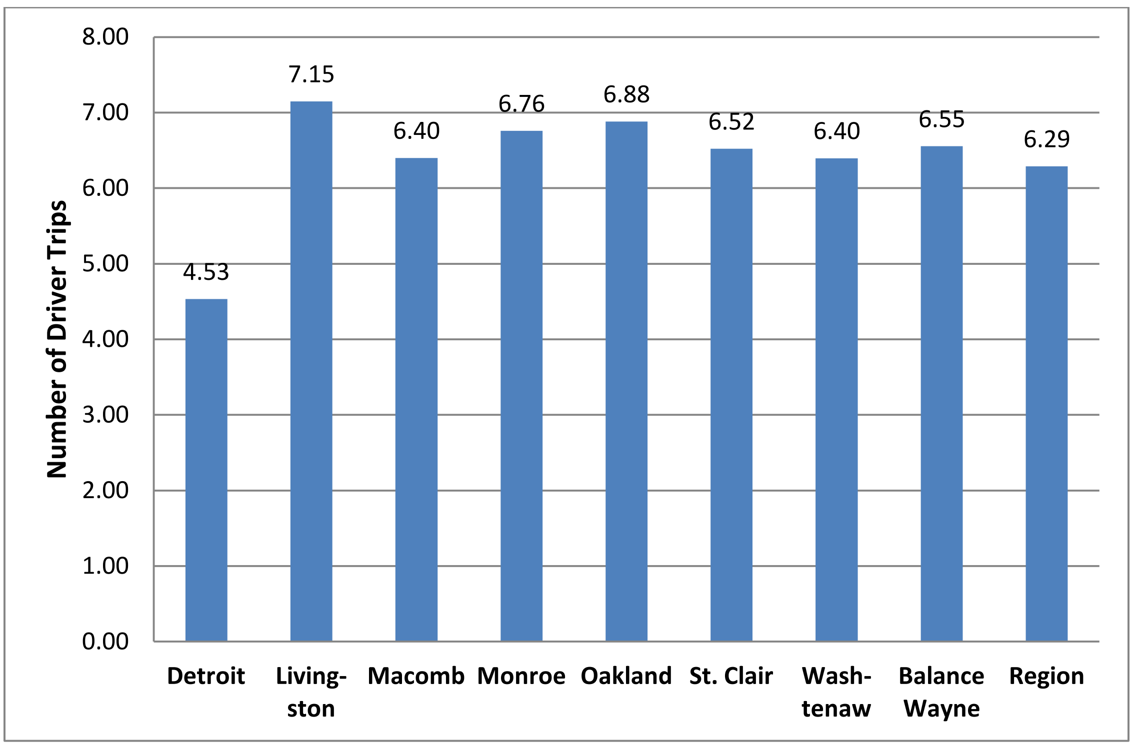
Income	Detroit	Living -ston	Macomb	Monroe	Oakland	St. Clair	Wash-tenaw	Rest of Wayne	Region
Less than $10,000	15%	2%	4%	3%	3%	5%	3%	4%	5%
$10,000 to $19,999	16%	5%	8%	8%	5%	13%	6%	8%	8%
$20,000 to $29,999	13%	7%	8%	13%	7%	14%	7%	10%	9%
$30,000 to $39,999	12%	7%	10%	8%	7%	10%	9%	9%	9%
$40,000 to $49,999	9%	7%	10%	10%	8%	9%	7%	10%	9%
$50,000 to $59,999	9%	7%	10%	11%	9%	10%	7%	10%	10%
$60,000 to $74,999	8%	10%	10%	10%	11%	15%	13%	12%	11%
$75,000 to$99,999	6%	18%	17%	15%	15%	10%	16%	13%	13%
$100,000 to $124,999	4%	16%	6%	6%	10%	4%	11%	7%	8%
$125,000 or more	2%	8%	4%	4%	12%	3%	11%	6%	7%
Under $50,000	2%	1%	2%	0%	2%	2%	1%	3%	2%
50,000 and over	1%	4%	4%	6%	5%	1%	4%	3%	3%
Don't Know	1%	1%	1%	2%	1%	1%	1%	1%	1%
Refused	2%	6%	6%	5%	4%	3%	4%	4%	4%

e) Travel Behavior

Besides the data about household characteristics, there is rich information about trips that households made in the travel survey. Some analyses on these trips were presented in the previous two chapters (e.g., Table 1, Figure 1, 2, and 5). Additional analyses are summarized below.

On average, households in the region made 6.29 trips each day. These trips can be chained in tours, and are not necessarily home-based. Households in Livingston County made more trips than households in any other places. This is not surprising given that they also led in household size, income, number of workers, and number of vehicles available. Households in Detroit made significantly fewer trips than households in any other areas. In fact, Detroit was the only place where number of trips was lower than regional average (Figure 14).

Figure 14. Number of Driver Trips



Livingston County also led in overall trip length measured in minutes (Figure 15) and work trip length (Figure 16). Average length of work trips was significantly higher than average length of all trips.

Figure 15. Mean Travel Time for All Trips

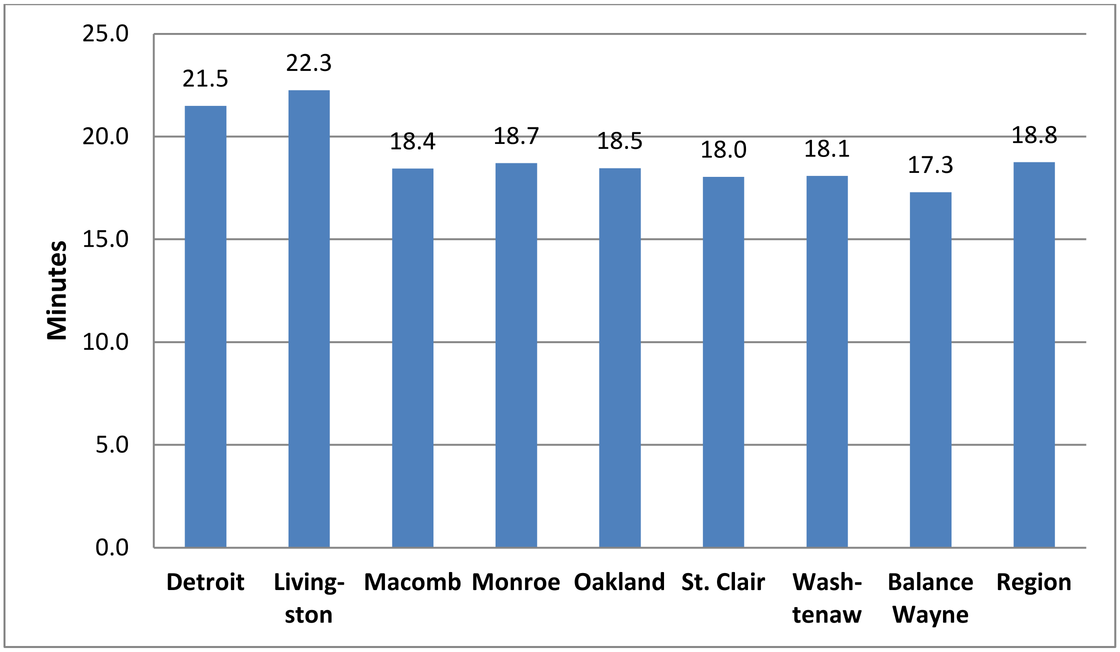
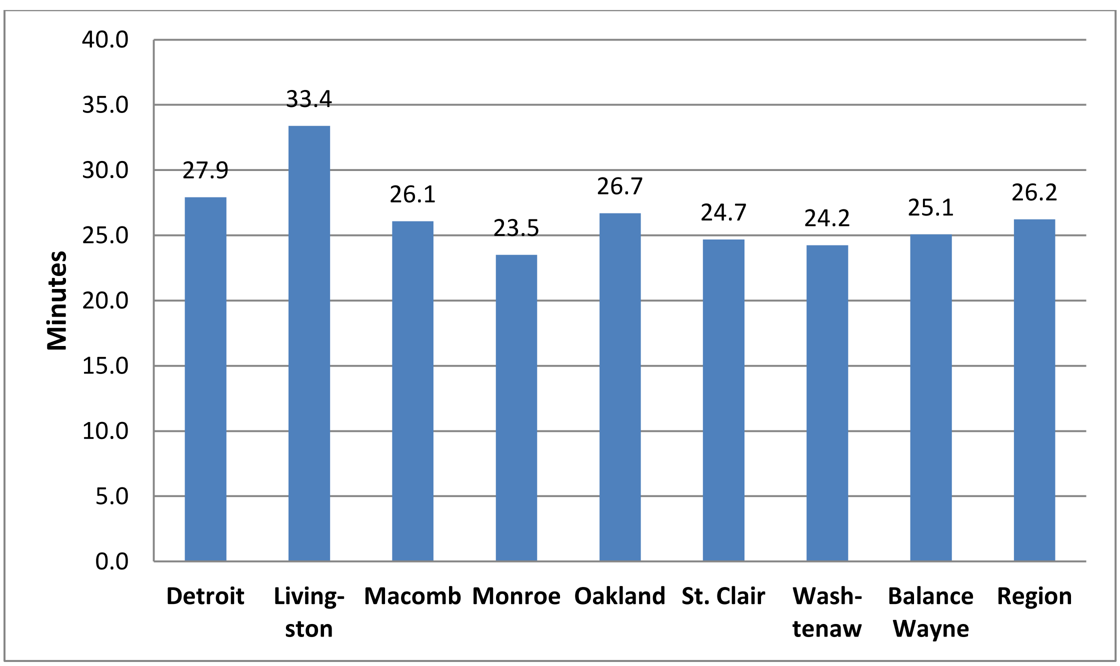


Figure 16. Mean Travel Time for Work Trips



It was more than ten years earlier when the previous household travel survey was taken in Southeast Michigan in 1994. Table 6 compares travel time by trip purpose from the two surveys. The comparison reveals that travel time for work trips increased somewhat. But travel time for nonwork trips decreased significantly. One might think that nonwork trips have become increasingly important in life and people attempt to minimize the cost of nonwork trips.

Table 6. Comparing Mean Travel Time by Trip Purpose, 1994 and 2005

Trip Purpose	1994 Survey	2005 Survey	Difference	Percent Difference
Home Based Work (HBW)	24.2	25	0.8	3.3%
Home Based Shopping (HBSH)	16.7	14	-2.7	-16.2%
Home Based School (HBSC)	17.2	11.8	-5.4	-31.4%
Home Based Other (HBO)	17.7	14.8	-2.9	-16.4%
Non-Home Based Work (NHBW)	19.4	20.8	1.4	7.2%
Non-Home Based Other (NHBO)	16.1	14.5	-1.6	-9.9%

3. Geocoding and GIS Processing

Survey consultants conducted initial geocoding to the sample households collected. That geocoding process was using street centerline files for geographical reference. Households and their trip stops were assigned latitudes and longitudes on street centerlines. For this study, parcels are the analytical unit. It is necessary to geocode households to parcels. The initial latitudes and longitudes from survey consultants were used as a reference only.

To geocode survey households to parcels, first the street address fields in parcel map and survey data are standardized. The fields used in this process included street number, prefix direction, prefix type, street name, street type, suffix direction, and ZIP code. Second, standardized addresses in survey data are geocoded to the addresses in parcel map using ArcMap GIS software. This automatic geocoding process resulted in 4,343 of total 6,065 households, or 71.6%, geocoded to parcels. Third, the remaining 1,714 households were geocoded manually. In the end, there were only eight households that could not be geocoded to parcels. The overall successful geocoding rate was 99.9% (Table 7).

Table 7. Geocoding Results of Survey Households

County	Total	Auto Geocoding	Manual Geocoding	Total Geocoded	Can't be Geocoded	Percent Geocoded
Livingston	440	310	130	440	0	100.0%
Macomb	966	663	302	965	1	99.9%
Monroe	348	192	156	348	0	100.0%
Oakland	1,228	953	271	1,224	4	99.7%
St. Clair	511	247	264	511	0	100.0%
Washtenaw	613	365	245	610	3	99.5%
Wayne	1,959	1,613	346	1,959	0	100.0%
Region Total	**6,065**	**4,343**	**1,714**	**6,057**	**8**	**99.9%**

GIS processing and analysis were used to link household attributes from household survey to the neighborhood characteristics for understanding the relationship between households and their locations to model household location preferences. Figure 17 shows one hypothetical (for confidentiality reasons) survey household over a digital ortho-photography. Also shown in the figure are land parcel boundaries in red. The analysis and modeling work described in this study were by parcel. The data obtained and used for characterizing these parcels will be explained in the following sections. There are over 1.8 million parcels in the study area.

Figure 17. A Sample Survey Household and Land Parcels

C. Parcel Map and Assessing Data

While the household travel survey data provide the attributes of the choosers for location choice models in this study, parcel data provide most of the characteristics of the choice set for the models. Parcel data consist of local land parcel maps and property assessment data associated with the land parcels. They provide highly disaggregated information and are the most logical form of source data for representing real estate characteristics in a disaggregated location choice model. Parcel data are increasingly available to researchers in the form of a GIS database due to rapid automation of land records. The parcel data used in this study were collected, processed, and enhanced by SEMCOG mainly for developing its regional forecasts.

1. Data Collection

Parcel data collection consisted of two primary data sets which were land parcel maps and property assessment records. In addition, local municipality master plans were collected separately but overlaid with the parcel data eventually. Furthermore, land use by parcel was added to the dataset, which was developed through the use of the digital parcel files, property assessment records, aerial photography, and other records about the buildings in the region such as CoStar commercial real estate database.

a) Parcel Maps

Digital parcel maps are typically created and maintained more centrally than assessment data. Most counties have high quality digital parcel maps but do not have the detailed assessment data as local assessors have, which are needed for the location choice analysis. This is partly because counties are mainly concerned with property assessing equalization and are mostly interested in assessing values only, whereas local assessors needed more detailed data for assessing the properties firsthand. While counties and municipalities have various schedules for updating their databases, the collected digital parcel maps and assessing data representing as closely as possible to the spring of 2005.

Parcel maps were collected from various sources but mostly from counties digitally. Livingston County, Macomb County, and Oakland County regularly update their digital parcel maps and provided their GIS layers closest to 2005. St. Clair County has been regularly updating its digital parcel layer but historically has not included the City of Port Huron's parcel data, therefore, City of Port Huron and the rest of county’s parcel maps were collected separately. Washtenaw County was similar to St. Clair County where City of Ann Arbor and the rest of county’s parcel maps were obtained separately.

Monroe County did not have a full county wide digital parcel file. SEMCOG staff, with assistance from Monroe County's Planning Department collected whatever digital parcel data existed in the county (both GIS data and CAD data) as well as paper maps. SEMCOG staff then processed all of the information and created a county wide digital parcel map for Monroe County.

The parcel map for Wayne County has not been updated since approximately 2000. The difference in dates between the parcel geography and assessment data caused significant mismatches. Extensive effort was taken at SEMCOG to help identify correct parcel to assessing record relationships. Meanwhile some municipalities in Wayne County have kept their own parcel files up-to-date. Their digital parcel maps, wherever available, were collected and incorporated into the Wayne County parcel file. These municipalities include City of Detroit, Brownstown Township, Canton Township, City of Dearborn, City of Livonia, Northville Township, City of Taylor, Van Buren Township, and City of Westland.

There are 1,837,676 parcels in the final parcel file for the Detroit region that are used for this study.

b) Property Assessment Records

While digital parcel maps depict the geographic boundaries of the land parcels, property assessment records provide the raw data for most of the location characteristics that are used as the input data for household location choice modeling. Property assessment records representing as close as possible to the spring of 2005 were collected, which was the assessment rolls archived after each municipality's 2005 March Board of Review (MBOR) process. Files containing 2005 MBOR assessment records were collected from municipality's assessors for the most of the region except for two small municipalities, one in Livingston County and another one in Washtenaw County, which never provided their assessing data. Their data had to be imputed using county data and data of similar parcels in the neighboring municipalities.

Local assessment files typically contain hundreds of fields. Some fields are always populated such as assessed values. Some other fields may not be populated in some municipalities. The most valuable attributes were extracted for modeling purposes as shown in Table 8. This table also included additional attributes that were not in the raw assessment data but were developed from other sources and added to parcels that could be used for modeling work.

Table 8. Key Attributes by Parcel

Name	Description
PRCLNUM	Parcel number that can be used to join parcel maps and assessing data
P_ADDRESS	Property address, including street number, name, city, state, and ZIP codes
TAXYEAR	Year of the assessed values
PROPCL	Parcel's current property class
EXEMPT	Exempt status. Can be used for identifying parcels owned by governments.
ASS1BLDG	Assessed improvement values, i.e., building values in the parcel
ASS1LAND	Assessed land value
YRBLT	Year when the building was built
SALEAMT	Amount of last sale
SALEDATE	Date of last sale
CIBLDGS	Number of commercial or industrial buildings
CIFLAREA	Commercial or industrial building square footage
RESBLDGS	Number of residential buildings
RESFLAREA	Residential building square footage
ZONING	Parcel zoning
Added Fields	
GENERAL_LU	General land use code assigned to the parcel
DETAIL_LU	Detailed land use code assigned to the parcel
STATUS	Development status assigned to the parcel, i.e., "developed", "undeveloped", or "undevelopable"
PLANYEAR	Year of the master plan updated
CATEGORY	Detailed planned use category as written in the community's master plan
MIN	Minimum planned density
MAX	Maximum planned density

2. Data Processing and Enhancement

Parcel-based assessment data were developed originally for local government taxation purposes. Therefore assessed value fields are most complete and accurate. Other fields that are less important for assessing purpose but are equally important to this study are less complete and less accurate than assessed values fields. Improvements to the raw assessing data were made in several ways for modeling purpose.

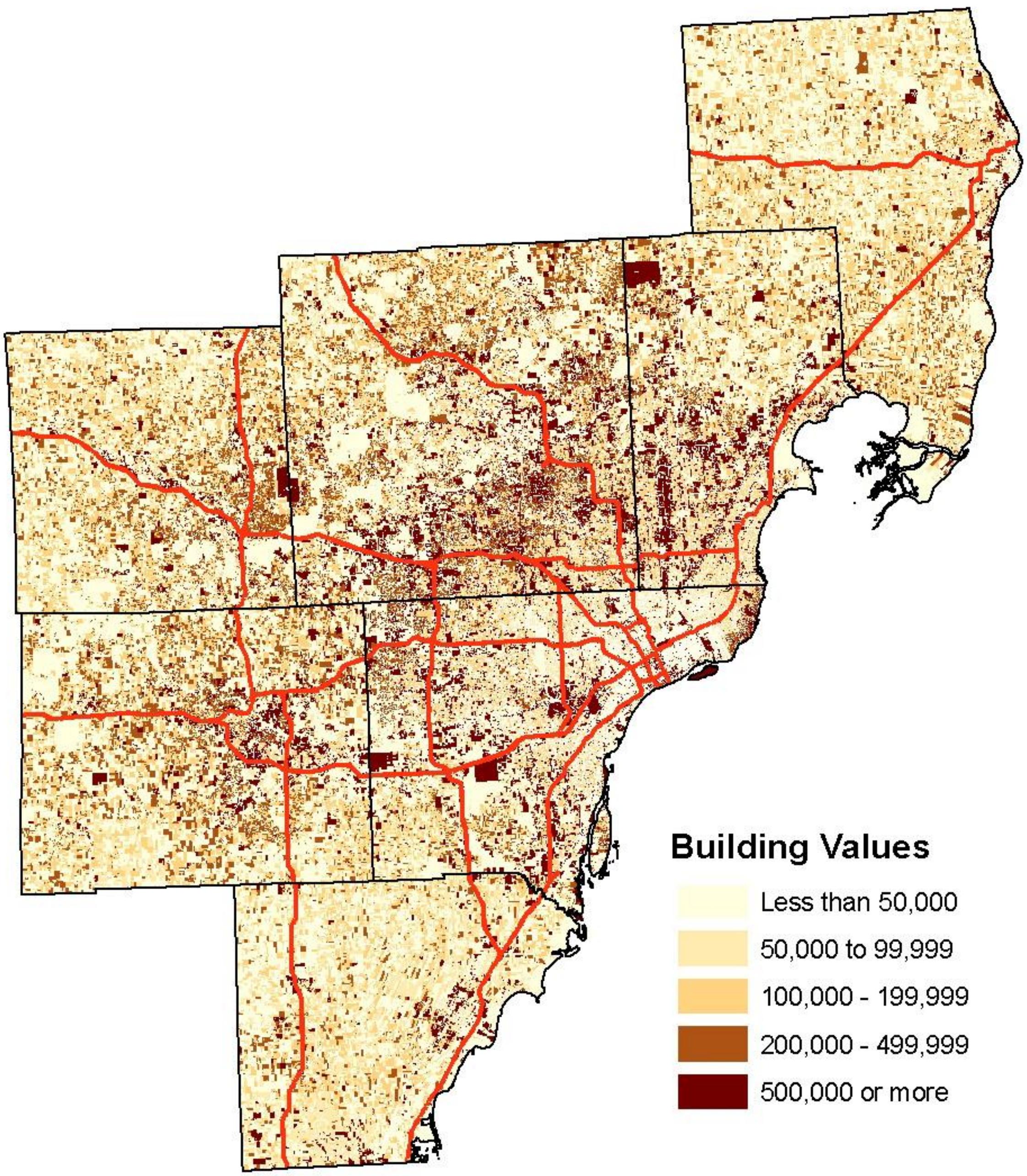
a) Imputing Missing Values

Even for assessed value fields, there are missing values, particularly for publicly owned parcels and buildings that are tax exempt. It is necessary to populate these missing fields as accurately as possible, based on valid data values and the relationships between attributes. The imputation process took several steps as summarized below.

Once the properties with missing values were identified, they were grouped by types of municipalities, including core city, small town, urban fringe, suburban, exurban, bedroom community, and rural area. Then a correlation analysis was conducted between missing fields and the fields for which data existed in the same type of municipalities. The next step was to utilize the results from the correlation analysis to develop regression models, relating the dependent variable (missing value) with the independent values (relevant variables). Finally, a value was generated for each missing data point using the regression models.

The result of this work was a complete layer of parcel map with necessary attributes for modeling location choice. For example, assessed land values and improvement values, i.e. building values, were assigned to every legitimate parcel. These values were used to represent the real estate prices of various locations in the region for modeling purposes. As Figure 18 shows, these values vary significantly across the region.

Figure 18. Assessed Building Values by Parcel



b) Completing Number of Housing Units

The quality of housing unit data varies in assessing data as some assessors are not particular interested in this type of data, particularly for apartments and mobile homes. An apartment database was created at SEMCOG by manually locating apartment buildings through assessment data as well as from other sources including calling apartment managements. The database contains records for all the apartment buildings found across the Detroit region. The database fields included number of units within each apartment building. Similarly, a manufactured housing database was created by manually locating manufactured housing parks in the region. The database contains records for all the manufactured housing parks in the region, and includes the field of total number of units within each manufactured housing park. Housing unit data in both apartment database and manufactured housing park database were used to update number of housing unit field of the parcel file.

Housing unit data by census block from 2000 Census and building permits data from 2001 to 2005 were also used to check the number of housing units in the parcel file. Necessary adjustments were made when significant mismatches were found.

c) Improving Land Use Classification

Land use types were developed through the use of information from the digital parcel files, property assessment records, aerial imagery, internet research, and other real estate data for the region. The list of parcel-based land use types is shown in Table 9.

Table 9. Land Use by Parcel

ID	Land Use Type	Description
1	agricultural	Agricultural Property
2	agricultural_residential	Property Used for both Agriculture and Residence
3	agricultural_mining	Mining Property
4	sf_residential	Single-Family Residential
5	mf_residential	Multi-Family Residential (Apartment, etc.)
6	mobile_home_park	Mobile home in a park
7	office	Commercial, Industrial, or Government offices
8	retail	Commercial Retail
9	entertainment_leisure	Commercial Entertainment/Leisure
10	hotel_motel	Hotels and Motels
11	commercial_mixed_use	Commercial Mixed Use
12	manufacturing	Industrial Manufacturing
13	warehousing	Industrial Warehousing
14	tcu	Transportation, Communication, Utility
15	hospitals	Hospitals
16	medical_facilities	Medical Buildings, Doctors Offices, etc.
17	public_building	Public Buildings (Library, Museum, etc.)
18	civic_organization	Veterans of Foreign Wars, Boy/Girl Scouts, etc.
19	schools	All Schools including Colleges and Universities
20	religious	Religious Organization Owned Property
21	parking	Parking Structures and Lots
22	park_open_space_conservation	Parks and Recreation/Conservation
23	ROW	Right of Way
24	vacant_developable	Vacant Developable Property
25	vacant_undevelopable	Vacant Undevelopable property
26	water	Water

d) Mapping Other Data to Parcels

Other data that were not originally parcel-based were converted to parcels. Master plans from municipalities are still mostly not parcel based at the present time. However, master plans collected in 2005 by SEMCOG were converted to parcels. Some of the fields in this file are listed in Table 8 above as “added fields,” including planned land use types and planned densities.

Digital parcel maps and property assessment data were joined by unique parcel IDs. The attributes in the parcel data can be used to compute additional variables to characterize neighborhoods for analyzing location choice. For example, linking household income to assessed value may reveal housing affordability, whereas analyzing age of buildings in an area may tell the maturity of the area and quality of housing that could affect household location choice.

Parcel maps and assessment data provided most of the data needed for characterizing the choice set but not all the data needed. Several other data sources provided additional information as discussed in the following sections.

D. Synthesizing Households

Household travel survey provides the sample data for characterizing the choosers in estimating household location choice models. Parcel data are used to measure many characteristics of the choice sets and the neighborhoods from which sample households can choose. What is missing from these two datasets is the characteristics of the other people that are not in the limited sample of household travel survey but live in the region. Who else are in the neighborhood can impact location choice significantly. Some studies use geographically aggregated demographic data for modeling purposes, such as census tract or TAZs. This study uses land parcel as the unit of analysis. It is necessary to develop a demographic data set at parcel level. Essentially, a universe of households and population is needed to represent the demographic patterns of the region in a highly disaggregated fashion. There are benefits of having such a universe of households and population, besides the requirement of parcel-based modeling. First, it can take advantage of accurate data from small geographies, for example 100 percent count of population by census block and housing value by parcel. Furthermore, individual households by parcel or building can be conveniently grouped into customized neighborhoods, e.g., areas within walking distance of a sample household.

Since there is no data source that provides individual households at such a detailed level directly, a mathematical procedure was used to “synthesize” them. A computer program for synthesizing households and population was pioneered by the Los Alamos National Lab for the TRANSIM software. The procedure uses several census data products and a series of Monte Carlo simulations to synthesize demographic characteristics for each household. This study used a similar method, but enhanced it by weighting more on marginal distributions at the smallest geography possible for each variable, and by considering local housing characteristics developed from parcel-based assessment data when assigning synthesized households to parcels. In the end, each one of the approximately 1.85 million households in Detroit region was synthesized and placed to parcels.

The synthesized households have seven attributes (Table 10) that are similar to the household attributes obtained from household travel survey. The one exception is that race of the head of household is also in the synthesized households but not in the household travel survey. Four major race categories are included, which are non-Hispanic White, non-Hispanic Black, Hispanic, and all others (mostly Asian).

Table 10. Attributes of Synthesized Households

Household Attributes	Values
Household Size	1, 2, 3, 4, 5, 6, 7 or more
Age of Household Head	Single year, 15 to 93 or older
Race of Household Head	White, Black, Hispanic, Other
Household Income	$0-$894,082
Number of Children	0, 1, 2, 3 or more
Number of Workers	0, 1, 2, 3, 4, 5 or more
Number of Vehicles Available	0, 1, 2, 3, 3, 4, 5, 6 or more

Three sets of census data were used in this process. First, for basic demographic information collected in decennial census ("short from" for Census 2000 and earlier), 100-percent data were available down to the block level in census Summary File 1 (SF1). Second, American Community Survey (ACS, or "long form" for Census 2000 and earlier) provides additional socio-economic sample data down to the block-group level. Third, individual household sample data are available in PUMS (Public Use Microdata Sample) for large geographic areas, called PUMAs, that have at least 100,000 people in each area. The objective of synthesizing households and population is to take the sample of individuals in PUMS that has the most detailed information to the much smaller geographies of block-groups to create a universe of individual households and people and ensure the characteristics of these households and people match what are observed at small geography such as block-groups, and eventually assigned to parcels.

Given that 100-percent cross-tabulated data are available at the block level for the key life cycle household characteristics of household tenure, household type, sex of householder, and broad age of householder, the Census 2000 Summary File 1 (SF1) table is chosen as the primary Census 2000 table to begin with. Additional variables are added to this from Census 2000 Summary File 3 (SF3). The basic idea is to use the more accurate 100-percent count data before using less accurate sample data.

There are three major steps in the household synthesis process as discussed below. The overall strategy is to prioritize the use of data sources by their coverage of household characteristics and synthesize characteristics that are consistent with aggregate block group level data while preserving multivariable distributions as represented by census PUMS data. Placement of the synthesized households into individual parcels is carried out using a ranked comparison of housing values and rental costs of the synthesized households to assessed property values in a digital parcel map.

1. Block-group Marginal Distributions

The first step is to produce marginal distributions of household characteristics at the census block-group level. This task develops a target distribution to guide the household synthesis process, ensuring that the characteristics of the households synthesized to each block-group will sum to the totals of the block-group for those characteristics from the Census Short Form, Long Form or American Community Survey (ACS) data. 100-percent count data are used first before sample data are used.

2. Drawing Households from PUMS

The second step is assigning weighted household records from PUMS areas to census block-groups. This is a simulation process. Weighted household records from the most recent PUMS dataset are expanded to form a complete universe of households in each PUMS geographic area (a.k.a. PUMAs). Each household is given a random number from a uniform distribution and then the dataset resorted, so that the drawing of households for assignment to block-groups is random. Each household is then drawn in turn and compared to each block-group within the PUMS area, in sequence. If a block-group can accommodate a household with the characteristics of the drawn household, then the household is coded to that block-group and removed from the draw. The marginal distribution of the block-group is decremented based on the household's characteristics, and the process repeats until each household has been assigned to a block-group, and each block-group has the appropriate number of households, whose characteristics sum to the block-group marginal distribution totals.

3. Assigning Households to Parcels

Once the synthesis of households has been completed, the PUMS households previously assigned to a census block-group are then placed to land parcels within the block-group. Land parcels containing housing units are identified using assessment data and other administrative records, as well as current land use and aerial photography layers. Census occupancy rates by tenure are applied to estimate total occupied housing units within a block group. Because there are differences between Census 2000 and parcel data, there may not be an exact match in the number of housing units and/or households at the block group level. However, by controlling the number of housing units assigned at the parcel-level to Census 2000 block group totals, the total number of households assigned to parcels can also be controlled to Census 2000 data at the block group level. In special circumstances where significant geocoding errors are apparent in Census 2000 data, block groups were aggregated together for the purposes of controlling to Census 2000. Placement of the synthesized households into individual parcels is carried out using a ranked comparison of housing values or rental costs of the synthesized households to assessed property values in the digital parcel file.

This household synthesizing method utilized the best data available from various sources. The placement of synthesized households at the parcel-level allows for the analysis of household data below traffic analysis zones and census blocks. Yet, individual households can always be analyzed at larger geographies by applying aggregation procedures.

The synthesizing process was run only once. But additional runs would not make much difference, for several reasons. First of all, this procedure uses the marginal distributions of socio-economic attributes of each census block-group to control synthesized households within a block-group. Census block-groups are small geographies, and variations of household characteristics within block-groups are small. In addition, assigning households to parcels within each block-group based on housing value further reduces uncertainty. The assignment process was carried out by a ranked comparison of the synthesized housing values or rental costs of households and the observed assessed values by parcel. The rank orders are fixed unless there are identical housing values, or identical rental costs, or identical assessed values in a block-group. Therefore, additional runs of the synthesizing procedure would not really make much difference.

E. Employment Data

An accurate and detailed employment dataset is always difficult to obtain. This study uses a published high quality 2005 QCEW (Quarterly Census of Employment and Wages) dataset by TAZ with some imputation.

1. Employment Data Overview

As described in the sections above, decennial census and estimations between census years are the best sources for estimating population and households. However, it is much more complicated for measuring employment than population partly because jobs can be measured in various ways, such as full-time jobs, part-time jobs, wage and salary jobs, self-employed or proprietary jobs, private jobs, and government jobs. QCEW is a virtual census of employment in the United States, covering over 99% of wage and salary civilian employment. The name QCEW was adopted in 2003. Before 2003, it was called Covered Employment and Wages program, because of its focus on jobs covered by Unemployment Insurance and Unemployment Coverage for Federal Employees. The QCEW program is often referred to as the “ES-202” program, as it is derived from an obsolete transmittal with that number that was part of the Employment Security (i.e., Unemployment Insurance) program (U.S. Department of Labor, Bureau of Labor Statistics, 2011).

QCEW is a cooperative program among the U.S. Department of Labor's Bureau of Labor Statistics (BLS) and the employment security agencies of the 50 States, the District of Columbia, Puerto Rico, and the Virgin Islands. In Michigan, State

Unemployment Insurance Agency collects and compiles employment and wage data for workers covered by the Michigan Employment Security Act. Among the data collected are monthly employment and quarterly wage information for workers covered by State unemployment insurance (UI) laws and for civilian workers covered by the program of Unemployment Compensation for Federal Employees (UCFE). Employment data represent the number of covered workers who worked during, or received pay for, the pay period including the 12th of the month. Excluded are members of the armed forces, the self-employed, proprietors, domestic workers, unpaid family workers, and railroad workers covered by the railroad unemployment insurance system. About 40% of agriculture wage and salary jobs are included in the dataset. Wages represent total compensation paid during the calendar quarter, regardless of when services were performed.

2. Employment Data Processing

Data from the QCEW program serve as an input to Federal and State programs, as well as research and analysis programs at federal, state and local level. Employment data are included in the QCEW program for nearly every industry by 4-digit NAICS (North American Industrial Classification System) code. The broad coverage, continuity, and currency of the QCEW program make it one of the most useful employment data sources for socio-economic research.

While QCEW data are readily available for large area analysis such as at national and state levels, there are several data quality issues and disclosure issues that must be addressed if the data are applied to small area analysis, such as at TAZ level. SEMCOG obtained March 2005 QCEW file form the State of Michigan, processed the data, and published a TAZ dataset for the Detroit region after confidentiality checks. Data processing and improvement procedures are summarized below.

a) Breaking out multi-establishment records

An employer may operate in a number of different locations, but there is often only one unemployment insurance account for an employer in QCEW data. For example, the U.S. Post Office employs thousands of individuals in hundreds of locations in Michigan. In fact, this applies to a number of large companies (e.g., General Motors Company), institutions (e.g., University of Michigan), and franchises (e.g., Domino's Pizza) in the region. Each location is called an establishment. To ensure the accuracy of employment data by small area, those records with multi-establishments are broken out by research based on company records or real estate and land use data.

b) Adjusting employment by type

In addition to multi-establishment records, firms may have various types of operations at the same location. A company may have significant numbers of research and development jobs and manufacturing jobs at one location. Some firms do not distinguish employment among various types when reporting to state unemployment insurance agency. Corrections and adjustments were made to deal with significant problems of this nature to provide more accurate measurement of the economy.

c) Geocoding

Employer addresses in QCEW and additional establishment addresses from the results of breaking out multi-establishment records were geocoded to parcel maps. Because of the complexity of the parcel addresses and QCEW addresses, there are several steps to geocode the establishments.

(1) The addresses were first geocoded to the parcel by using the original parcel addresses. But the matching rate is only about 30%.

(2) Many large parcels, such as malls, offices, etc, have address ranges, but the parcel map only contain a single address. A script was used to create address ranges for those parcels. The script uses the addresses adjacent to a large parcel to create address ranges. For example, if an address of a large parcel is 2 Maple Rd, and the address of the adjacent parcel is 10 Maple Rd, the script assigns the address ranges, 2 to 8 to the parcel. The addresses that were not geocoded in step one were run through these created address ranges. The matching rate increased significantly to approximately 70%.

(3) Next, the remaining addresses were geocoded using Michigan Geographic Framework (MGF). The points were then moved to appropriate parcels based on parcel maps and aerial photos.

(4) Finally, remaining addresses were geocoded individually by using mapquest.com, aerial photos, and other sources.

(5) A small number of addresses that could not be geocoded were allocated to the TAZs, proportional to geocoded jobs in TAZs.

(6) The University of Michigan and Eastern Michigan University addresses were manually geocoded to the buildings by using aerial photos.

Employment records geocoded to parcels were summed to TAZs.

d) Confidentiality requirements

To assure the anonymity of QCEW covered firms, public disclosure of all QCEW data is contingent on the number of Unemployment Insurance covered accounts (i.e., firms or employers) included in an individual data record (e.g., TAZ employment), and each account's share of the employment reported in that data record. Any publications based on data from the QCEW program can not disclose data for any level in which the universe (1) consists of fewer than three unemployment insurance accounts; or (2) is dominated by a single employer that represents 80 percent or more of employment. A computer program was developed at SEMCOG to check the processed data for confidentiality restrictions. As the result, 2,453 of total 2,811 TAZs passed total employment confidentiality checking, which represents approximately 92% of total jobs in the region (Table 11). Employment by type was also checked using the same computer program. Once the disclosure restrictions were met, the employment dataset was approved by the State of Michigan, and then published at SEMCOG Web site.

Table 11. Results of Confidentiality Checking on Employment Data

	Number of TAZs	Percent	Number of Jobs	Percent
Total	2,811	100.0%	2,044,057	100.0%
Published	2,453	87.3%	1,876,944	91.8%
Blocked	358	12.7%	167,113	8.2%

e) Imputation of missing data

Although the blocked 358 TAZs (or 12.7%) are mostly small zones with little employment, it is desirable to have a complete employment dataset for all the TAZs. Therefore, the missing 8.2% employment (or 167,113) in those blocked TAZs was imputed based on the published employment data and land use data.

First, for TAZs in which total employment numbers were blocked, total blocked jobs in each county are allocated to these TAZs based on employment land area from the region's land use layer. This is possible because no county had either total employment numbers or employment numbers by type were blocked. The assumption of this method is that jobs are distributed evenly and proportional to employment land use, including commercial, industrial, governmental and other institutional land areas.

Once the missing total employment numbers for blocked TAZs were imputed, the blocked employment numbers by industrial class were imputed for each blocked TAZ by the following steps.

- Calculate total blocked jobs by industrial class by county.
- Distribute above calculated blocked jobs in each industrial class in each county to TAZs that had jobs blocked in that industrial class in proportion to total jobs of those TAZs.
- Apply iterative-proportional-fitting method to initially imputed jobs (and imputed jobs only) by industry to match TAZ total blocked jobs and county blocked jobs by industrial class. Note that, in this application, iterative-proportional-fitting is a mathematical method that iteratively adjusts cell values of a two-dimensional matrix so that sum of each column or row matches a predefined total. In this case

the predefined columns or rows are TAZ total blocked employment numbers and county blocked employment numbers by industrial class respectively.

The result of the imputation work is an employment dataset of 2,811 TAZs with 2,044,057 jobs by 16 industrial sectors. This is the employment dataset used for this research.

F. Summary of Chapter III

This study is data intensive and relies on input data from a wide range of sources. The disaggregated nature of the input data for this study allows for investigations of accessibility at great detail and at various scales of geography. It also makes it possible to study individual household's residential location choice that takes into account fine-grained variations of socio-economic, land use, and transportation characteristics in their neighborhoods.

The input data discussed in this chapter are developed to characterize the "choosers" and the "choice sets" for estimating household location choice models. For characterizing the "choosers," 6,057 households from the most recent household travel survey are selected and processed. For characterizing the "choice set," data were developed for over 1.8 million land parcels in the study area. Assessing data and synthesized households were all located to the land parcels. Employment data are also used to measure accessibility at TAZ level.

The next chapter will discuss the methodologies used in this study. This will include methods of measuring accessibility at local, mid-range, to regional scales;

methods of weighting accessibility with the degree of clustering of activities; as well as developing residential location choice models.

CHAPTER IV

METHODOLOGY

This chapter describes methods used in analyzing the effects of various aspects of accessibility on residential location choice in Detroit region. This study develops a number of accessibility measures and uses these measures as independent variables in multinomial logit models of residential location choice. The study uses these models to assess the importance of accessibility by estimating and interpreting the relative influence of accessibility measures on residential location choice.

This chapter first addresses a number of modeling issues in order to apply the discrete choice theory to this study. These issues include creating a statistically representative sample from household travel survey, and sampling of housing units from a large number of housing units in the entire region to construct a practical choice set.

This chapter then describes in detail the methods for analyzing various aspects of accessibility and developing specific measures, including three components of measuring local accessibility and the composite local accessibility index developed from these components, regional accessibility by vehicle availability, mid-range accessibility at a number of scales, spatial clustering of destinations measured by either single nearest neighbor method, i.e., Average Nearest Neighbor (ANN) spatial statistics, or by multi-distance spatial cluster analysis, i.e., Ripley's K-function, as well as individual commute time to work for up to two workers in each household.

This chapter also discusses the techniques and advantages of using multinomial logit models for analyzing residential location choice, selections of independent variables including accessibility measures, and the expected results of estimating such models in terms of the signs of the independent variables.

A. Multinomial Logit Model for Estimating Residential Location Choice

This study uses discrete choice modeling techniques to estimate residential location choice, particularly the effects of accessibility on the choice. Discrete choice models are widely used in many areas, including travel demand modeling in the urban and regional planning field. It was Daniel McFadden's Nobel-prize winning work on Random Utility Theory and his derivation of the generalized extreme value class of models, which include multinomial and nested logit models, that gave these models a firm foundation within econometrics (McFadden 1973, 1981, 1984). Discrete choice models have since become standard methods in developing models that attempt to predict individual choices among a finite set of alternatives. Discrete choice models are generic in the sense that they do not impose overly-restrictive assumptions on the choice process, and have been shown capable of addressing large and complex choice sets effectively (Ben-Akiva and Lerman 1985). The purpose of model estimation is to reveal the coefficients that describe the strength of each chosen variable in affecting the choice.

The “results” are the coefficients, interpreted to give a sense of what factors are most important in determining the choice.

The discrete choice approach models an individual’s selection of a single choice from a number of alternative choices. In this approach, a model can use explicitly both attributes of the individual who makes a choice decision and the characteristics of the alternative choices from which the individual selects. The word “discrete” indicates another property of the discrete choice approach, that is an individual selects from a limited number of available choices. The idea is that the individual chooses his or her optimal alternative from all available choices, given a budget constraint. However that individual chooser needs not necessarily to optimize all the attributes of all the factors. Both properties of the discrete choice approach make it suitable for this study of analyzing residential location choice. First, both the attributes of house hunters and the characteristics of locations affect decision making when searching for a place to live. Second, a locating household selects a particular location not by being able to optimize all factors but in fact by selecting a bundled package deal that may offer the best combination of affordable price, local services, neighborhood amenity, as well as accessibility at various scales, among other factors.

For the discrete choice models specified and estimated in this study, the choosers are households obtained from the 2004-05 Detroit region household travel survey. The choice set for each chooser consists of the chosen property, which is the house where the household actually lives, and a number of randomly selected other properties in the region, a.k.a. the rejected properties. The total number of housing units in the choice set for each household is 30 in this study. The purpose of the modeling work is to analyze how accessibility affects households' decision on where to live, controlling for other factors. The dependent variable is the choice of a location, whereas the independent variables are household attributes and location characteristics including accessibility measures. Household attributes include age, race, income, number of persons, number of children, number of workers, and number of motor vehicles available in the household. Location characteristics include assessed land value, assessed improvement value (value of buildings), residential density, mix of housing units and commercial space measured by square feet, crime rates, school quality, accessibility, and others that will be discussed later in this chapter.

The assumption underpinning the discrete choice approach to residential location choice modeling is that observed residential location patterns are the results of choices made by individuals. A rational consumer selects the alternative from all possible alternatives that maximizes his or her utility. Each alternative has a utility function for that consumer. The attributes of the alternatives may be positive or negative as expressed by the signs of the independent variables in the utility function equations. The attributes are bundled. The consumer chooses the alternative that maximizes utility from all available alternatives.

Expressed in mathematical terms, a consumer chooses alternative *i* if:

$U (X_{in}) > U (X_{jn})$ for all alternatives *j* in choice set C

where

U = utility

X_i = the chosen alternative

X_j = other alternatives

n = the chooser making choice decisions

C = the choice set of all alternatives for the chooser to choose from

The utility function has attributes of the chooser and characteristics of the alternatives. The chooser's attributes can only enter the utility equation when they interact with the characteristics of the alternatives. The utility function has a deterministic component which is the function of observed attributes, and an error term which represents the unobserved factors. An analyst's knowledge of the chooser's and alternatives' utility function is only partial. And a chooser may never have all the information to make choices. Therefore, one may observe that seemingly similar choosers facing apparently same choice sets may be making different choice decisions. As a result, models cannot make deterministic predictions about individual choices. Instead, models predict probabilities of individual choices. Nevertheless, the probability of a choice being selected increases when the deterministic component of the utility function to an individual chooser increases, or when the deterministic components of the utility of other alternatives to that individual decrease. Mathematically, the probability that the utility of the chosen alternative *i* exceeds the utilities of all other alternatives *j* for an individual equals the following:

$$P_n(i) = \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})}$$

Where

$P_n(i)$ = probability that chooser *n* selects alternative *i*

V = the deterministic component of utility function

C_n = the choice set

This form of discrete choice model is called the multinomial logit model (Domencich and McFadden 1975; Ben-Akiva and Lerman 1979). The most commonly used method for estimating multinomial logit model is the maximum likelihood estimation (Greene 2003). Maximum likelihood estimation first constructs a likelihood function equivalent to the probability of the observed sample. It then uses a gradient search method to determine the value of the estimated coefficients so that the logarithm of the value of the function is maximized.

The model estimation results can be reviewed and evaluated in a number of ways. First, the direction (positive or negative signs) and significance of the estimated parameters can tell if independent variables, such as accessibility measures at multiple geographic levels in this study, are important in household location decision making process. Secondly, relative influence of variables can be estimated and compared. Finally, the explanatory power of multiple models can be compared by assessing the goodness of fit measurements.

When applying discrete choice approach to this study, two technical issues need to be dealt with first as described below, before actually constructing the models.

1. Reducing the Choice Set to a Practical Size

This study estimates multinomial logit models for household location choice using residential properties as the choice set. Ideally, the form of the model is appropriate for estimating choice among the full set of alternatives available to households if it is practical in terms of data collection and statistical analysis at a very detailed and disaggregated level. The study area for this research has over 1.9 million housing units. All the housing units are identified at parcel level in a database as described in Chapter III. However, it is infeasible to estimate multinomial logit models using such a large choice set. Considering the enumeration of all the alternatives is impractical when the universal choice set is too large for multinomial logit model estimation, McFadden (1978) studied and proved that reduced choice sets from the full universe could yield consistent parameter estimations. He considered three ways of reducing the full choice set (C) to a manageable subset (D) of alternatives as follows:

a) choose a fixed subset D of C, independent of the observed choice,

b) choose a random subset D of C, independent of the observed choice, and

c) choose a subset D of C, consisting of the observed choice and one or more other alternatives, selected randomly.

McFadden proved that in all three types of subsets, consistent estimates of the parameters of the utility function can be obtained from a fixed or random sample of alternatives from the full choice set, as long as the multinomial logit functional form is valid. Therefore, analysis of housing location can be carried out with a limited number of alternatives, facilitating data collection and processing, provided the choice process is described by a valid multinomial logit model.

In this study, the third type of subset is used to reduce the choice set for practical reasons. A random subset of all housing units is selected which includes the chosen housing unit and a number of alternatives which are considered as the rejected housing units. I tested various sizes of the randomly selected subset, including 30, 50, and 200.

The results show consistent estimates of the parameters. The final models presented in this dissertation use 30 choices.

2. Adjustments to Household Travel Survey Sample

When SEMCOG conducted household travel survey in 2004 and 2005, it over-sampled households in smaller and often more rural counties than in larger and more urban counties. Table 12 compares each county's share of households from 2010 U.S. Census to those from household travel survey. All four smaller counties, which are Livingston, Monroe, St. Clair, and Washtenaw Counties, were significantly over-sampled, whereas the three larger counties, i.e. Macomb, Oakland, and Wayne Counties, were under-sampled. It was necessary to create a statistically representative sample before estimating models for this study.

Table 12. Deviation of Sample Size from Census Household Distribution

	Census Households		Household Survey		Difference	Reduced
	Number	Share	Number	Share	in Share	Sample
Region	1,845,218	100.0%	5,720	100.0%	0.0%	4,224
Livingston County	55,384	3.0%	387	6.8%	3.8%	127
Macomb County	309,203	16.8%	896	15.7%	-1.1%	708
Monroe County	53,772	2.9%	310	5.4%	2.5%	123
Oakland County	471,115	25.5%	1,133	19.8%	-5.7%	1,079
St. Clair County	62,072	3.4%	476	8.3%	5.0%	142
Washtenaw County	125,232	6.8%	588	10.3%	3.5%	287
Wayne County	768,440	41.6%	1,930	33.7%	-7.9%	1,759
City of Detroit	336,428	18.2%	941	16.5%	-1.8%	770
Balance of Wayne county	432,012	23.4%	989	17.3%	-6.1%	989

One way to correct over-sampling in small counties is to reduce number of sample households in these counties so that the distribution of households in each county will be the same as census data. The reduced sample size for each county is shown in the last column of Table 12.

Another way to correct the sampling problem is to assign weights to samples in various counties, so that the weighted samples are proportional to the true probability of selection. Mathematically, each observation is assigned a weight that equals to the ratio of its county's true share of census household numbers to the county's share of households in household travel survey. Model estimation will then be based on the weighted samples, which are equivalent to the samples shown in the last column in Table 13. Manski and Lerman (1977) developed a model estimator, called the WESML (Weighted Exogenous Sample Maximum Likelihood) estimator. It can use the weighted samples for estimating multinomial logit models.

Table 13. Derivation of Sample Weights by County

	Census Households		Household Survey		Weighting	Adjusted
	Number	Share	Number	Share	Factor	Sample
Region	1,845,218	100.0%	5,720	100.0%	1.00	5,720
Livingston County	55,384	3.0%	387	6.8%	0.44	172
Macomb County	309,203	16.8%	896	15.7%	1.07	958
Monroe County	53,772	2.9%	310	5.4%	0.54	167
Oakland County	471,115	25.5%	1,133	19.8%	1.29	1,460
St. Clair County	62,072	3.4%	476	8.3%	0.40	192
Washtenaw County	125,232	6.8%	588	10.3%	0.66	388
Wayne County	768,440	41.6%	1,930	33.7%	1.23	2,382
City of Detroit	336,428	18.2%	941	16.5%	1.11	1,043
Balance of Wayne county	432,012	23.4%	989	17.3%	1.35	1,339

This present study uses the sample reduction method to create a statistically representative sample for model estimation, because the sample weighting method had not been implemented in the version of the model estimation software, UrbanSim, that was used for this study.

B. Model Specifications – Independent Variables

This section provides a list of the independent variables considered in the residential location choice models, with brief explanations. The methods of developing the key accessibility variables are further explained in the following sections.

1. Affordability/Economic Variables

"income_-_cost": This variable is the income of a household interacting with the price of housing, defined as the natural log of household income minus annualized housing cost. It measures the affordability of housing relative to income, which arguably represents one of the most important factors that influence a household's choice of housing, because housing price should be a powerful indicator of a household's location choice given the income of the household.

"housing_size_x_high_inc": This variable and the following two variables measure the preference and ability to pay for various sizes of housing. This variable is defined as households with an annual income equal to or greater than $75,000 interacting with square footage of housing units.

“housing_size_x_mid_inc”: Households with annual income equal to or greater than $25,000 and less than $75,000 interacting with square footage of housing units.

“housing_size_x_low_inc”: Households with annual income less than $25,000 interacting with square footage of housing units.

2. Social Composition Variables

“race1_x_zonal_pct_race1”: This variable and the following three variables represent similarities of households in a neighborhood in regard to race. This variable is defined as a household whose head is Non-Hispanic White interacting with percent of households whose head is Non-Hispanic White in the Traffic Analysis Zone (TAZ).

“race2_x_zonal_pct_race2”: A household whose head is Non-Hispanic Black interacting with percent of households whose head is Non-Hispanic Black in the TAZ.

“race3_x_zonal_pct_race3”: A household whose head is Hispanic interacting with percent of households whose head is Hispanic in the TAZ.

“race4_x_zonal_pct_race4”: A household whose head is any race other than the above three interacting with percent of households whose head is any race other than the above three in the TAZ. The majority of these households are Asian in Detroit region, but also included in this group are Native Americans, Alaskan Natives, and multi racial households.

“high_inc_x_zonal_pct_high_inc”: This variable and the next variable represent similarities of households in a neighborhood in regard to income. This variable is defined as a household with annual income equal to or greater than $75,000 interacting with percent of households of the same range of income in the TAZ.

"low_inc_x_zonal_pct_low_inc": Households with annual income less than $25,000 interacting with percent of households of the same range of income in the TAZ.

3. Amenity and Service Variables:

"school_quality": Average scores from Michigan Educational Assessment Program (MEAP) standardized testing for grade 3 to 8, by Intermediate School District (ISD), 2010.

"school_quality_x_children": Households that have children interacting with average MEAP scores.

"school_quality_x_high_inc": Households with annual income equal to or greater than $75,000 interacting with average MEAP scores.

"crime_rate_violent": Number of violent crimes per 100,000 residents, by municipality, 2010.

"crime_rate_property": Number of property crimes per 100,000 residents, by municipality, 2010.

"crime_rate_both": Number of violent crimes and property crimes per 100,000 residents, by municipality, 2010.

"crime_rate_both_x_high_inc": Households with annual income equal to or greater than $75,000 interacting with number of violent crimes and property crimes per 100,000 residents, by municipality, 2010.

"property_tax": Property_tax rate by municipality, 2010.

"is_detroit": A dummy variable, equaling 1 for City of Detroit and 0 for all other areas.

4. Accessibility Variables

Accessibility variables are listed below. The following Section C of this chapter explains the methods for measuring these accessibility variables in greater detail.

“local_access_density”: Population density within walking distance, which is a circle of a quarter of mile radius.

“local_access_diversity”: Square feet of commercial buildings within walking distance.

“local_access_design”: Number of 4-way or more intersections within walking distance.

“local_acess_composite”: An index created by using factor analysis that combines the above three local accessibility measures.

“local_acess_young_hh”: The composite local accessibility measure interacting with households whose head is 30 years old or younger.

“mid_range_access”: Mid-range accessibility measured by number of jobs in retail, education, leisure, and other services within 10 minutes of travel time.

“emp_access_x_workers>cars”: Logsum-based regional accessibility to employment interacting with households that have more workers than vehicles available.

“emp_access_x_workers<=cars”: Logsum-based regional accessibility to employment interacting with households that have the same or more vehicles than workers.

“worker1_commute_time”: The first worker’s travel time from home to his/her employment zone.

“worker2_commute_time”: The second worker’s travel time from home to his/her employment zone.

“cluster_k”: Spatial cluster index defined by using multi-distance spatial cluster analysis (a.k.a., Ripley’s K-function).

“cluster_k_x_emp”: Ripley’s K-function based cluster index interacting with employment.

While the above discussions grouped potential independent variables by subject, one may also categorize these variables by geographic scales. When a household considers its residential location, it may consider characteristics of the house specifically, attributes of the neighborhood, and its relationship to the rest of the region. In such a classification scheme, the first set of location attributes is about the housing itself. These attributes include the cost of housing that can be measured by the assessed value of the property in this study. The cost of housing and its relationship to household income is a vital variable theoretically for residential location choice. Housing type also affects location choice. Large families with more children may be more likely to choose single family homes, whereas young households without children or senior empty-nesters may prefer multifamily housing. Other housing specific variables include the size of a house and when a house was built.

The second set of variables is about the immediate neighborhood. This is where local accessibility being measured and incorporated in this study. These variables correspond to the activities that can be reached by walking or other non-motorized mode, over a distance of approximately a quarter of a mile. Being able to walk to daily-life destinations may be attractive to at least some households. The local accessibility measures are developed by spatial queries of a house's neighboring parcels. Achieving this scale of analysis provided the possibility to model location choice and travel behavior at a level that can effectively represent pedestrian and bicycle scales of travel. As a result, it may provide a basis for making more systematic assessments of the effects of urban design-scale policies on both location choice and travel behavior. Traditional zone-based travel models are severely limited by poor performance on intra-zonal travel and insufficient representation of non-motorized travel modes. By creating a more detailed basis for the land use model, a major barrier to the improvement of transportation planning to address non-motorized modes and the integration of urban design policies might be effectively removed.

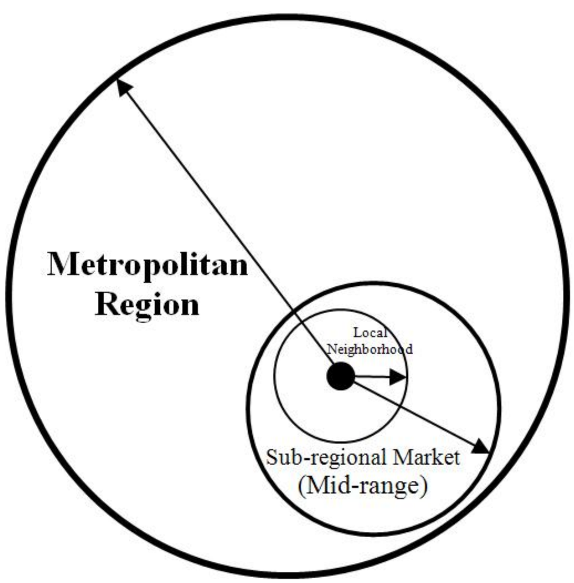
The third set of variables deal with characteristics of the larger neighborhood, including a wide range of attributes such as who else lives in the neighborhood, what are the shopping options, school quality, and crime rates. These variables reflect sub-regional market areas. Mid-range accessibility proposed by this study belongs to this group. For example, number of retail and commercial square feet that are mostly reached by nonwork trips is a mid-range accessibility variable. Definition of the "mid-range" is tested empirically with various distance and travel time thresholds in the present study. The task of the modeling work is to determine if the mid-range accessibility has independent effects on location choice controlling for local and regional accessibility.

The last set of variables deal with the regional context. Travel time from residential areas to downtown, employment centers, airports, and other regional destinations may be relevant to a household's location choice. This study uses logsum-based regional accessibility to represent the regional context for residential locations, and focuses on if clustering of destinations has significant effects on household location choice. Spatial clustering based weights are included in the models to test if more clustered destinations have greater impact on accessibility and residential location choice. The next section discusses accessibility measures at various scales.

C. Measuring Multiple Aspects of Accessibility

This study categorizes accessibility into metropolitan region, sub-regional market areas or mid-range, and local neighborhood levels as shown in Figure 19. Regional accessibility deals with how to access employment opportunities in the entire region. Mid-range accessibility focuses on the ease to access nonwork destinations such as shopping and daily-life service activities. Local accessibility measures opportunities for accessing destinations by non-motorized model.

Figure 19. Three Scales of Accessibility



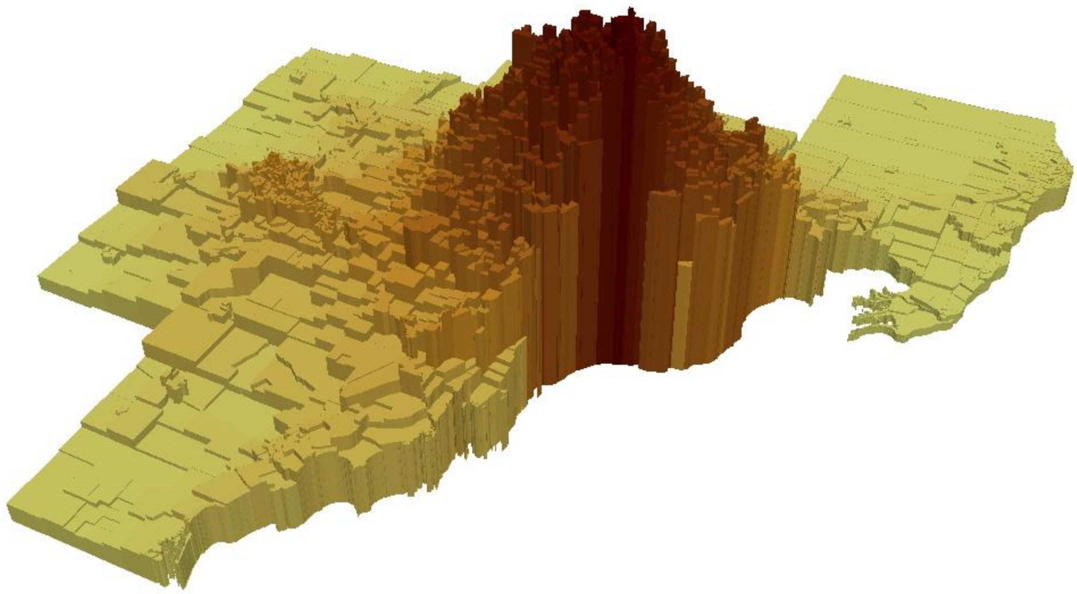
1. Regional Accessibility

This study measures regional accessibility using a utility based method that takes into account distributions of employment opportunities in the region and the impedance affecting access to them. The impedance is represented by logsum from SEMCOG's travel demand forecast model. Mathematically, the logsum is the denominator of the mode choice model in the travel demand forecast modeling system.

SEMCOG has two types of households in is mode choice model: 1) households that have fewer motor vehicles available than number of workers in these households, and 2) households that have the same number or more vehicles available than number of workers. A separate set of logsum numbers is available for each type of households. Both household-type-specific logsums and combined logsum numbers can be used to measure regional accessibility. When using combined logsums, all households use the same set of logsums for calculating regional accessibility. When using household-type-specific logsums, households of each type uses their respective set of logsum numbers.

Regional accessibility to employment by TAZ calculation in this study was measured by using zone-to-zone logsums and number of jobs by TAZ. The concentration of high-regional-accessibility zones is obvious (Figure 20). City of Detroit and surrounding areas have much higher employment accessibility than other areas. This picture shows little effects of the decentralization of jobs. That is one reason why regional accessibility itself may not explain residential location choice very well.

Figure 20. Regional Accessibility to Jobs in Detroit Region



2. Mid-range Accessibility

This study views mid-range accessibility as the accessibility of sub-regional market areas within a metropolitan region. It focuses on measuring accessibility to nonwork daily life destinations. Specifically, the attraction of these destinations is measured by the employment numbers in four industries published by SEMCOG. They are retail trade, education, leisure, and other services. These industries are selected because they represent nonwork trip destinations of typical daily life. The selection of industries is also constrained by data availability. For example, the "other services" industry is one of the 16 industries published in SEMCOG's employment data by TAZ. It comprises establishments engaged in providing services not specifically defined elsewhere in the North American Industry Classification System (NAICS). Establishments in this sector are primarily engaged in such activities as providing drycleaning and laundry services, personal care services, death care services, pet care services, photofinishing services, temporary parking services, and dating services; promoting or administering religious activities, grantmaking, and advocacy. But it also includes establishments that provide commercial and industrial machinery and equipment repair and maintenance, which are not likely daily life destinations.

Mid-range accessibility supplements and balances regional "attractions" measured by regional accessibility and the "convenience" measured by local accessibility. Existing analyses of accessibility have shown high degrees of correlation of accessibility among various trip purposes (Srour et al. 2002). Places that are highly accessible for work trips are also highly accessible for shopping and social, recreational trips and vice versa. Yet, analyses on travel behavior have shown significantly greater willingness for people to travel longer distance for work than for nonwork purposes. This study measures work accessibility at the regional scale and nonwork accessibility at sub-regional scale.

There are conceptual and measurement challenges in defining accessibility at the sub-regional level. The present study uses a cumulative opportunity measure, which counts number of jobs indentified for representing daily life nonwork destinations within a certain travel time threshold. Definition of the mid-range travel time threshold was explored in a trial-and-error process. Travel behavior analysis (Figure 2 in Chapter I) indicates that nonwork trips as a percent of total trips declined to less than 50% once the distance was longer than 15 miles. Various ranges of sub-regional market areas were tested empirically beginning from a 5-minite travel time threshold up to 30 minutes by 5-minutes increment. Travel time is absent from the local accessibility measures in the present study, because they assume that speed does not vary significantly within walking distance. When distance to activities increases, it is necessary to consider travel time in the mid-range accessibility measures. It is arguably better to use travel time than distance because metrics that rely on proximity alone neglect the possibility that travel may be slowed by denser development.

Household location choice models were estimated with control variables and each of the mid-range accessibility measures at various time thresholds. The estimation results show that mid-range accessibility with 10-minute travel time threshold has the most significant effect on residential location choice. Therefore, the mid-range accessibility variable used in the final models is defined as number of jobs in retail, education, leisure, and other services within 10 minutes of travel time.

3. Local Accessibility

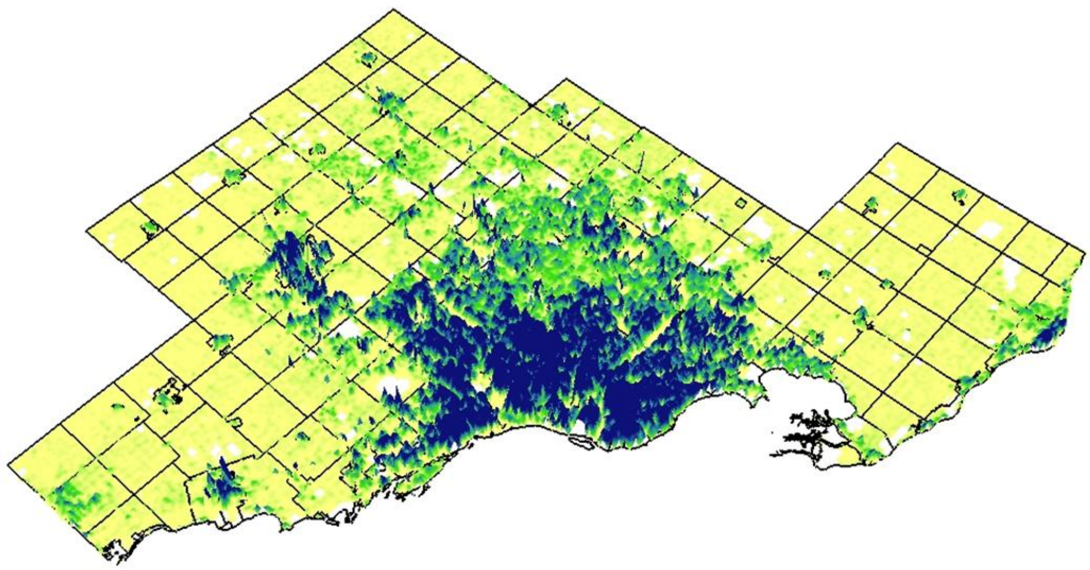
In this study, local accessibility focuses on measuring the effects of non-motorized modes at the neighborhood level. Specifically, local accessibility was measured by a composite index that takes into account of density, diversity, and design (3D) characteristics of the neighborhood (Cervero and Kockelman 1997). To measure local accessibility for each parcel, a neighborhood of a quarter-mile radius is defined around the centroid of each parcel, using the "Neighborhood Statistics" tool in ArcGIS software. Each parcel has four values calculated based on the characteristics of the neighborhood in regard to local accessibility. First, population density of the neighborhood is calculated to gauge the intensity of the neighborhood. Second, square footage of commercial buildings in the neighborhood is calculated to indicate the degree of land use mix. Third, number of four-way or more intersections was calculated in the neighborhood to represent traditional neighborhood design characteristics. Finally, a composite index based on those three measures was created using factor analysis to measure the overall local accessibility of the neighborhood surrounding a parcel.

The Neighborhood Statistics tool in ArcGIS is a function that rasterizes an input GIS coverage and computes an output raster where the value at each location is a function of the input cells in a specified neighborhood of the location. For each cell in the input raster, the Neighborhood Statistics function computes a statistic based on the value of the processing cell (i.e. the cell at the center) and the value of the cells within a specified neighborhood, then it sends this value to the corresponding cell location on the output raster (ESRI 2011). This study uses Neighborhood Statistics function to create a raster data set from the original vector parcel file. This raster data set is made up of 25 meter by 25 meter gridcells, and the neighborhood is set to be a circle of a quarter-mile radius around each cell. "The neighborhood" can be imagined as a "moving window" of a quarter-mile radius circle that moves across the entire region one cell at a time. At each move, the tool calculates a value based on the characteristics of all the cells that fall inside the window and assigns that value to the cell at the center of the circle.

a) Population Density

Population numbers were assigned to parcels in the process of synthesizing household and population. ArcGIS Neighborhood Statistics tool transfers population numbers from parcels to gridcells, and then calculates the sum of population of all cells in the neighborhood (i.e. within a quarter-mile radius circle) and assigns the sum to the cell at the center of the neighborhood. This sum of population represents the total population in the neighborhood that can be accessed by non-motorized mode. The higher value indicates better local accessibility (Figure 21). In this map, the higher points and the darker colors represent higher population density, which indicates greater local accessibility.

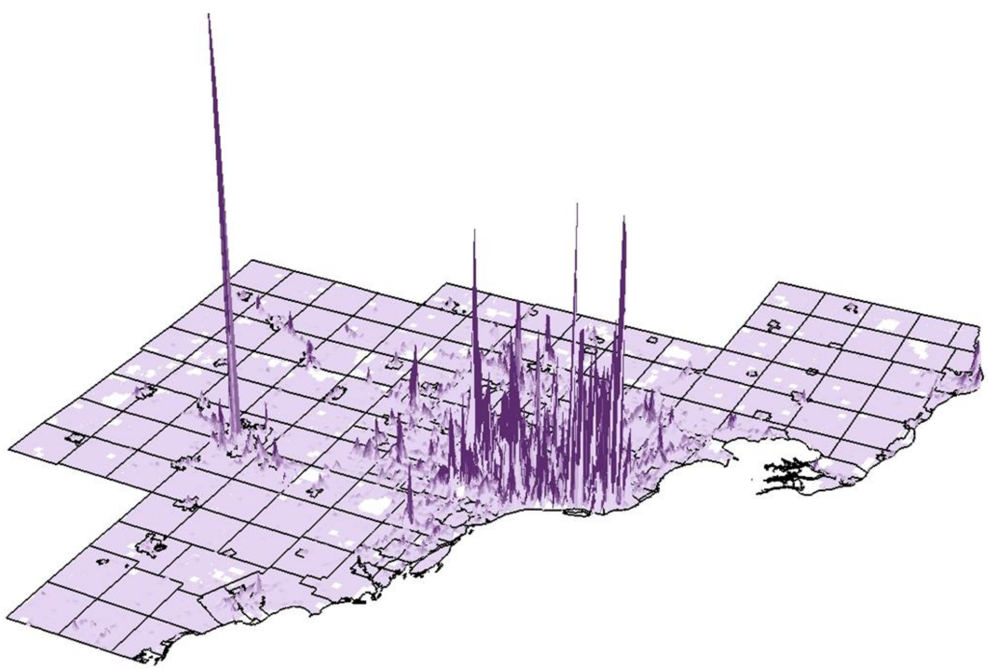
Figure 21. Local Accessibility: Population within Walking Distance (1/4 Mile Radius), Southeast Michigan, 2008



b) Diversity - Land Use Mix

Mixing of population with non-residential land use is critical to increase accessibility because non-residential uses provide meaningful destinations for the residents. This study uses square footage of commercial buildings to gauge land use mix. “Commercial buildings” in this study include retail, entertainment, and commercial services that could be attractions for non-motorized modes of travel such as walking and biking. These square footage data were developed from building types and actual building size information from parcel files. Similar to population density calculation, ArcGIS Neighborhood Statistics tool transfers square footage from parcels to gridcells, and then calculates the sum of square footage of all cells in the neighborhood and assigns the sum to the cell at the center of the neighborhood. This sum of square footage represents the commercial attractions in the neighborhood. Higher values mean better local accessibility, as shown in Figure 22, where higher points represent greater square feet of retail, entertainment and leisure, and commercial uses, indicating greater local accessibility.

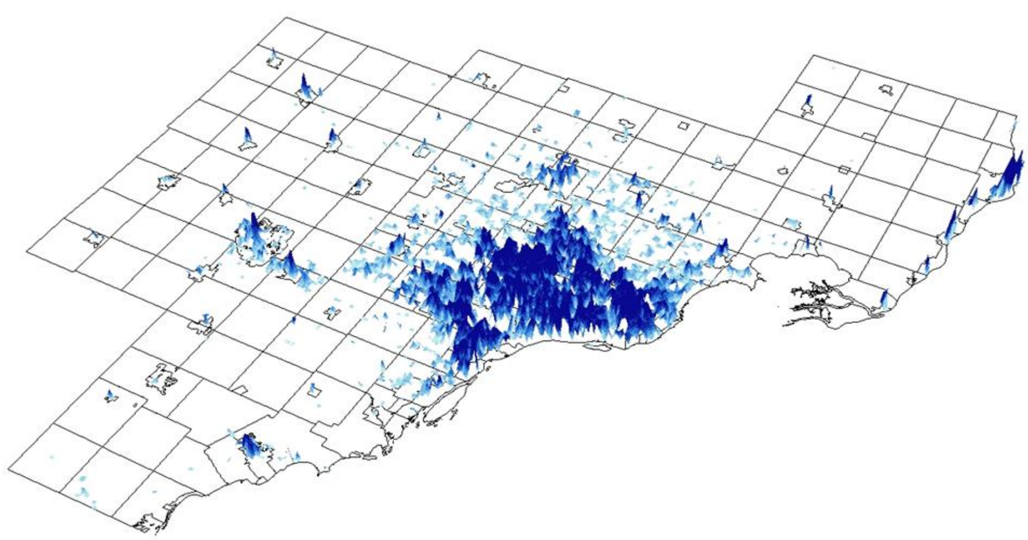
Figure 22. Local Accessibility: Commercial Square Footage within Walking Distance (1/4 Mile Radius), Southeast Michigan, 2008



c) Design

In regard to design features that affect local accessibility, this study uses number of four-way (or more) intersections to represent the traditional neighborhood characteristics. Contrary to modern subdivisions with cul-de-sacs, traditional grid road networks have been found to provide higher local accessibility, and counting number of four-way intersections is an effective method of distinguishing traditional neighborhoods from suburban subdivisions (McNally and Kulkarni 1997). Again, ArcGIS Neighborhood Statistics function creates a raster data set from the original four-way or more intersection points based on detailed road network. It then calculates the sum of such intersections in the neighborhood and assigns the sum to the cell at the center of the neighborhood. The higher sum of such intersections indicates enhanced connectivity of destinations, particularly for non-motorized mode such as walking in the neighborhood (Figure 23). High accessibility areas based on number of four-way intersections are shown by the higher points and darker color.

Figure 23. Local Accessibility: Number of 4-way (or greater) Intersections within Walking Distance (1/4 Mile Radius), Southeast Michigan, 2008



d) Composite Index of Local Accessibility

Density, diversity, and design may re-enforce each other or undermine each other's effects on accessibility in neighborhoods. A composite index based on all three local accessibility measures is developed for this study by using GIS overlay and statistical factor analysis technique.

First, a regional parcel-centroid point coverage was created from the parcel file. This parcel centroid coverage was then overlaid with the three raster output files from the above analysis of each local accessibility measure. ArcGIS Spatial Joins tool assigns the values from raster files to the parcel centroids. These values are population, commercial square footage, and number of four-way intersections. The next step was to combine those three values into a composite index for each parcel centroid, using factor analysis.

The main idea of factor analysis is to form a new variable, from a set of existing variables, that contain as much variability of the original data as possible. This is a method of data reduction. The reason to reduce the number of variables is to handle data more easily in the model, so that it is possible to use one variable to represent various aspects of local accessibility. This study used factor analysis function in SPSS to create the local accessibility index from the three individual local accessibility measures.

Local accessibility is based on 1.8 million parcels in the region. Local accessibility maps (Figure 21, 22, and 23) show different patterns than regional accessibility (Figure 20). The peak values of local accessibility are more prominent and more scattered throughout the region, whereas regional accessibility, based on 2,811 Traffic Analysis Zones (TAZs), shows more gradual changes from the region's center to the edges.

4. Commute Time – Individual Worker's Journey to Work

This study considers not only place-based accessibilities as described in the about sections, but also individual workers' journey to his or her own job, i.e., commute time. Workers' home and work locations were obtained from the SEMCOG household travel survey. While workers' home locations were geocoded to parcels, work locations were geocoded to Traffic Analysis Zones (TAZs). Travel time during morning peak hours from a worker's home location to his or her work TAZ was used to measure the worker's personal accessibility to employment.

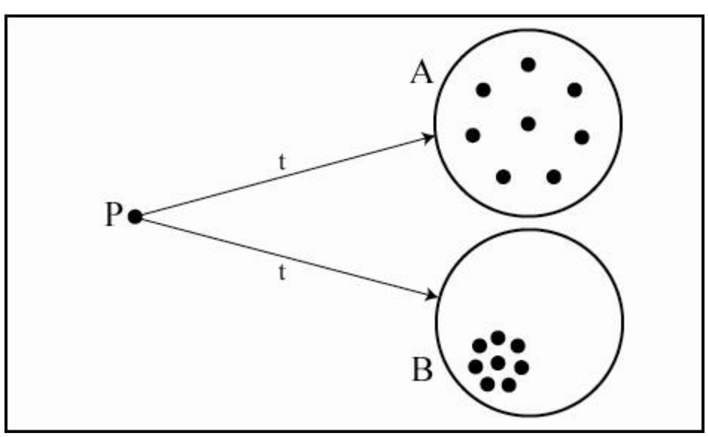
A worker's specific employment location may be a key factor that affects his or her residential location choice. When there are multiple workers in a household, multiple employment locations of these workers may all affect the household's residential location decision. This study considers up-to-two workers' journey to work in residential location choice models.

D. Spatial Clustering Analysis

Spatial cluster analyses help to measure the degree of clustering of destinations in the region. Two types of spatial clustering measures are tested in this study. First, a straightforward "Average Nearest Neighbor" (ANN) spatial statistics was used. This is a simple and easy to interpret measure, but it looks only at the nearest neighbor's distance. Other point-to-point distances are ignored. Second, a multi-distance spatial cluster analysis, i.e. "Ripley's K-function" spatial statistics was then explored. This measure addresses limitations of looking only at the nearest neighbors used by the average nearest neighbor method. It looks at the distribution of distances between all pairs of points. Therefore it is a more robust measure of spatial clustering. Both ANN scores and K-function scores are calculated at both municipality and TAZ levels. Once the spatial clustering is measured, the number of jobs, cluster scores, and the interaction between the two were tested in residential location choice models.

Number of jobs, their clustering, and the interaction of the two may all affect location choice (Figure 24). Using a conventional accessibility measure, e.g., gravity-based accessibility measure, person P's accessibility to zone A and zone B are the same because each zone has the same amount of activities, represented by the centroids of eight non-residential parcels (e.g., commercial, industrial, and institutional parcels) and the travel time to each zone being the same "t". But if P travels to more than one destination in chained trips of a tour, one may argue that accessibility to B is higher because destinations in B are close to each other which may reduce travel time for trips between these destinations. The task of cluster analysis is to measure the degree of clustering of destinations in these zones.

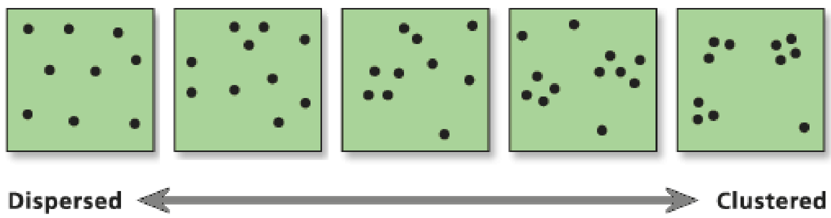
Figure 24. Attraction in zones of various degrees of clustering



1. Average Nearest Neighbor (ANN) Spatial Statistics

One method to measure spatial clustering is to use the Average Nearest Neighbor (ANN) spatial statistics as illustrated in Figure 25 (ESRI 2012).

Figure 25. Average Nearest Neighbor (Spatial Statistics)



The Average Nearest Neighbor method measures the distance between each feature centroid and the nearest neighbor's centroid. In this study, these are the centroids of non-residential parcels in zones. It then averages all these nearest neighbor distances. If the average distance is less than the average for a hypothetical random distribution, the distribution of the features being analyzed are considered clustered. If the average distance is greater than a hypothetical random distribution, the features are considered dispersed. The index is expressed as the ratio of the observed distance divided by the expected distance, while the expected distance is based on a hypothetical random distribution with the same number of features covering the same area. If the index is less than 1, the pattern exhibits clustering. If the index is greater than 1, the trend is toward dispersion. Statistical significance can be tested by Z scores. Z scores are measures of standard deviation. For example, if an analysis returns a Z score of +2.5, it is interpreted as “+2.5 standard deviations away from the mean.”

An example is shown in Figure 26 and Figure 27. Each dot in Figure 26 represents a commercial, industrial, or institutional parcel in the City of Birmingham, Michigan. The grey lines depict the city boundary. In this example, AAN ratio is 0.32 that is calculated as observed mean distance of the dots divided by their expected mean distance if randomly distributed, indicating a clustered pattern. $\boldsymbol{Z_{ANN}}$ score is calculated as -20.86 standard deviations. The critical $\boldsymbol{Z_{ANN}}$ score for a statically significant clustered pattern at 99% level is -2.58 standard deviations. These results mean that there is less than 1% likelihood that the clustered pattern of non-residential parcels in the City of Birmingham could be the result of random chance (Figure 27).

Figure 26. Non-residential Parcels in Birmingham, Michigan

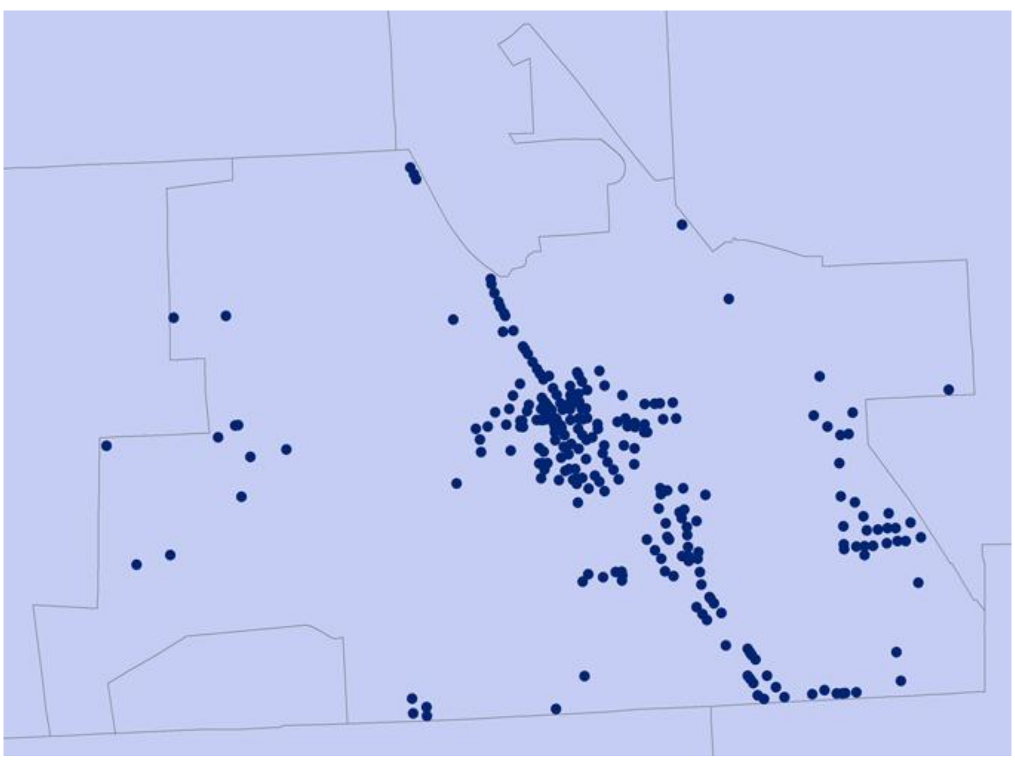
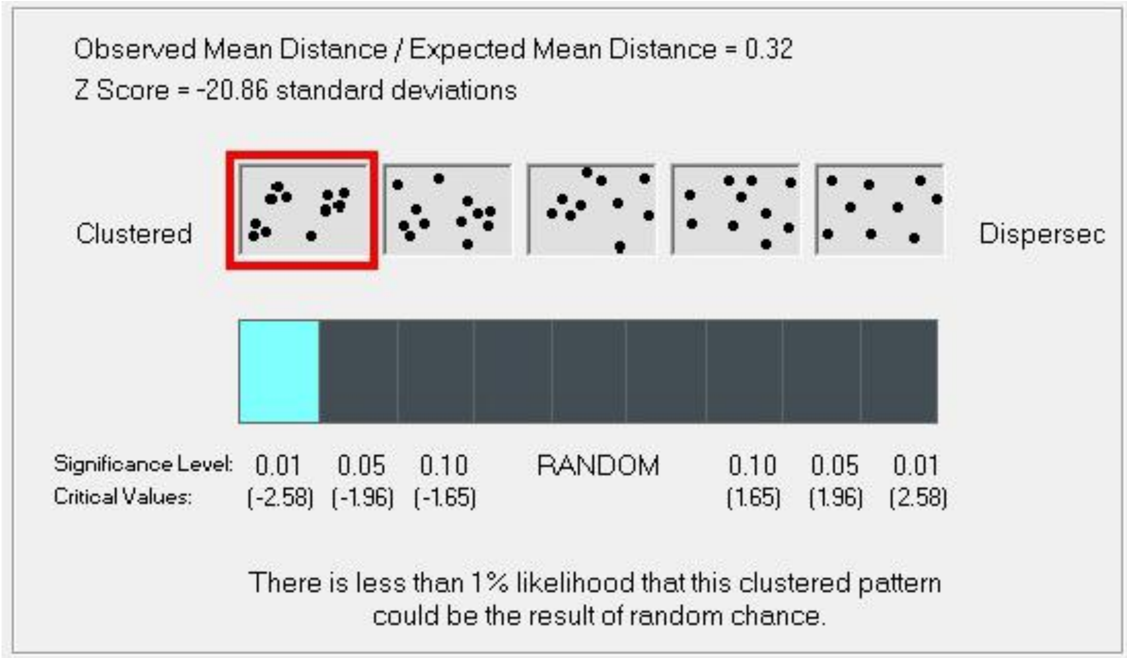
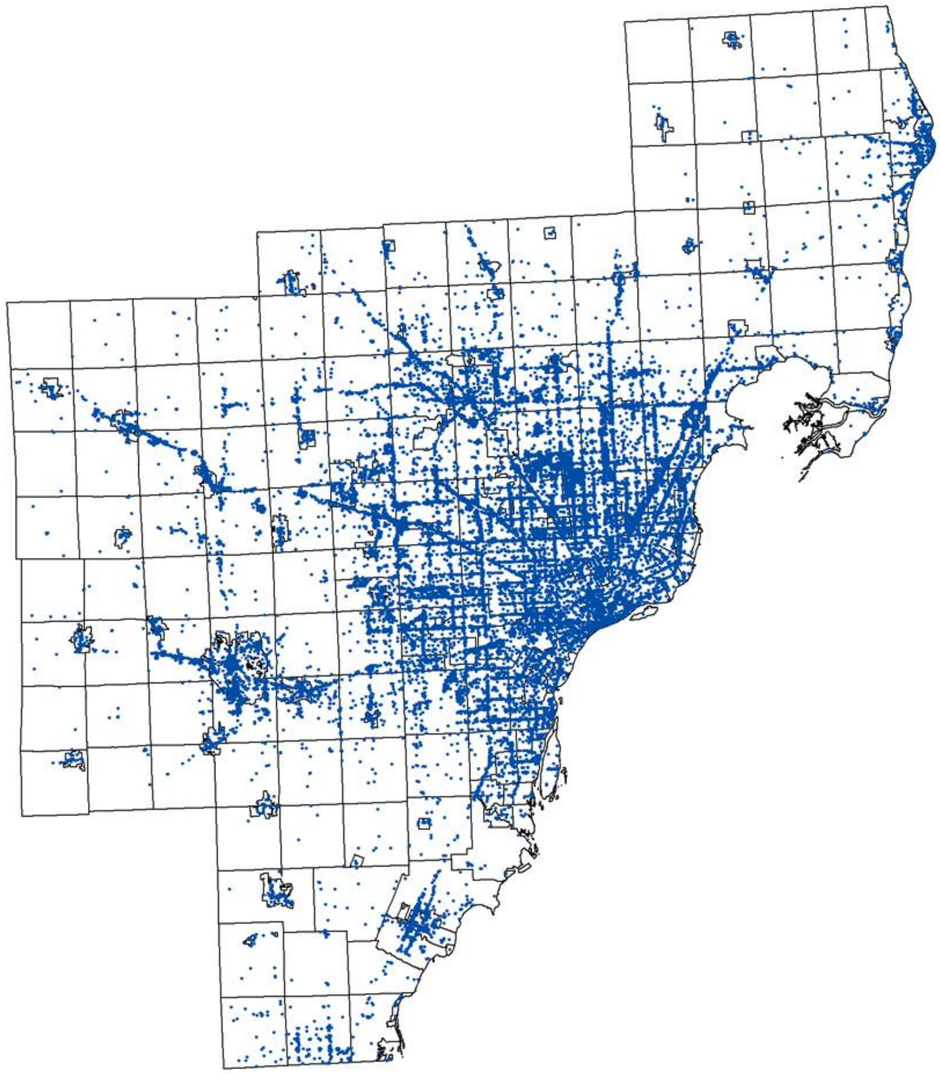


Figure 27. Average Nearest Neighbor Statistics for Birmingham, Michigan



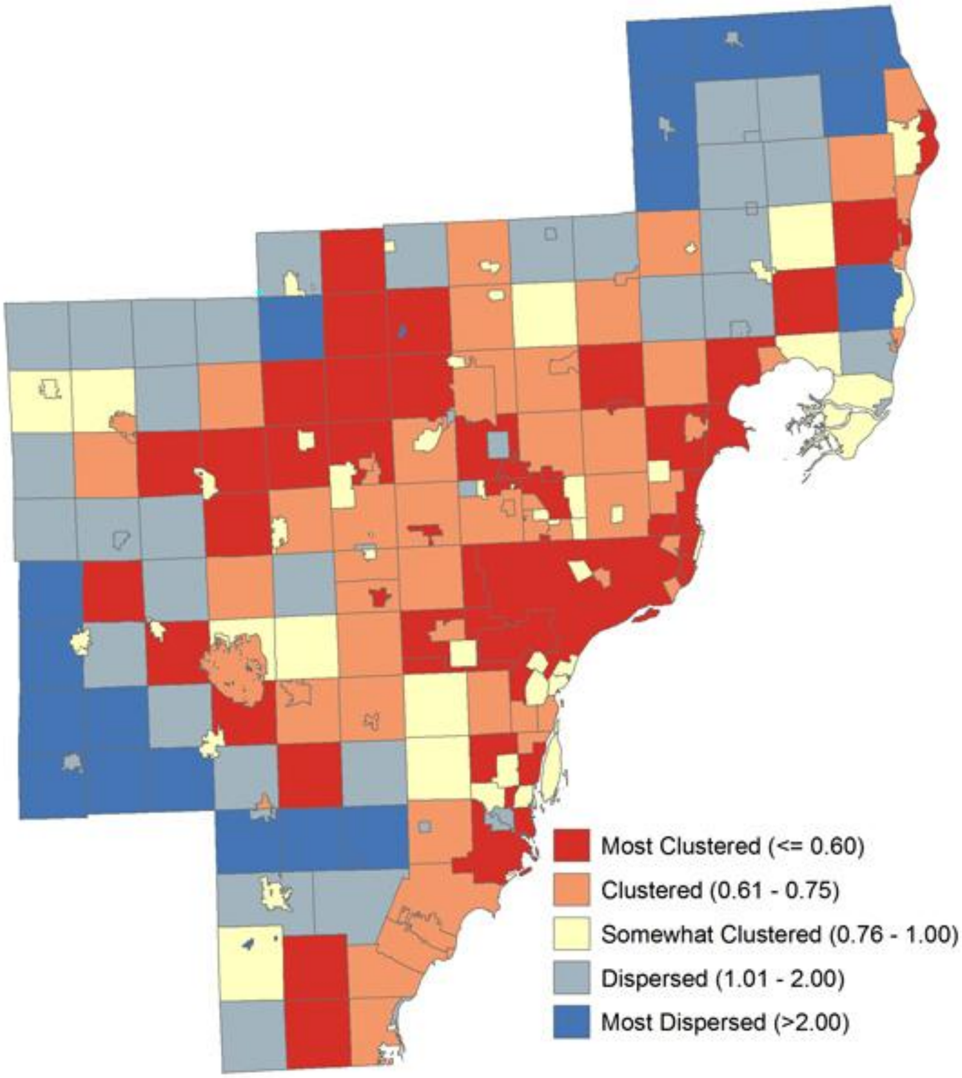
This study uses the centroids of all the parcels that have commercial, industrial, or institutional parcels to represent the destinations in the region. Each point in Figure 28 depicts such a destination, whereas the black lines depict the municipal boundaries. The Average Nearest Neighbor method analyzes the points in each municipality to calculate an index that measures the degree of clustering of the destinations in that municipality. ANN scores are also calculated by TAZ. Both sets of ANN scores by municipality and by TAZ are tested in the models, because cluster indices are often affected by the unit of analysis.

Figure 28. Destinations for Cluster Analysis



It is noteworthy that Figure 28 shows clearly the relationship between destinations and transportation networks. Most destinations cluster around major roadways in the region. Municipalities that have dense road network tend to have more clustered destinations. Average Nearest Neighbor scores of these destinations by municipality are shown in Figure 29.

Figure 29. ANN Scores by Municipality



Compared to regional accessibility to employment (Figure 20), cluster scores measured by the ANN ratio (Figure 29) show a more decentralized pattern. If regional accessibility is weighted by cluster scores, some areas that have fewer but more clustered jobs may have higher accessibility than non-weighted measures would show. This may be a desirable effect because the residential location pattern in the region is more decentralized than the non-weighted regional accessibility indicates.

2. Multi-Distance Spatial Cluster Analysis (Ripley's K-function)

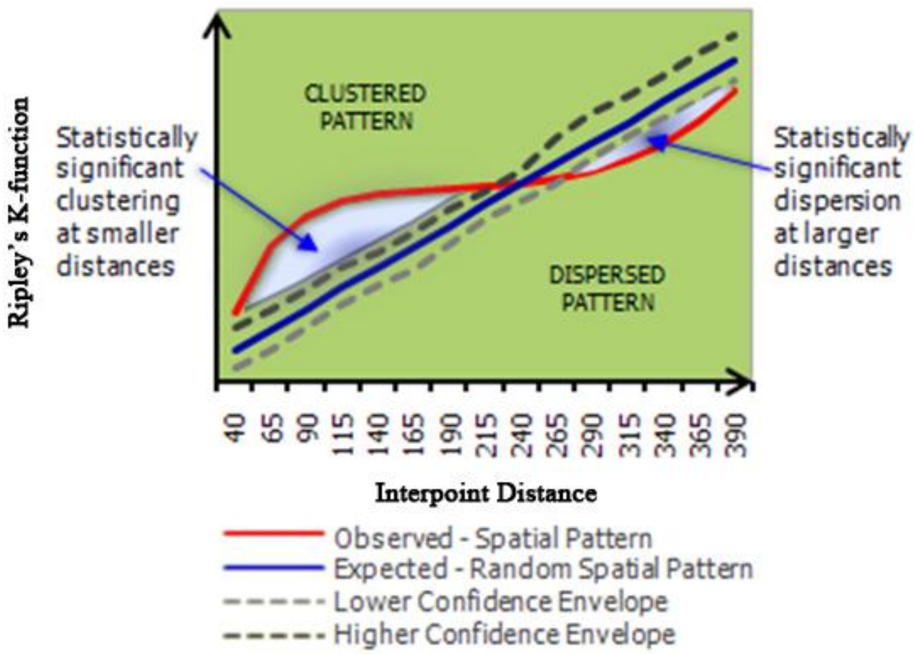
Another method to measure spatial clustering is to use a multi-distance spatial cluster analysis, i.e. "Ripley's K-function" spatial statistics (Ripley 1981). A distinguishing feature of this measure from the previous ANN measure is that it summarizes spatial dependence (feature clustering or feature dispersion) over a range of distances. When using this measure, the user must specify the number of distances to evaluate and, optionally, a starting distance and/or distance increment. With this information, the average number of neighboring features associated with each feature is computed. Neighboring features are those closer than the distance being evaluated. As the evaluation distance increases, each feature will typically have more neighbors. If the average number of neighbors for a particular evaluation distance is greater than the average concentration of features throughout the study area, the distribution is considered clustered at that distance. Mathematically, Ripley's K can be defined as the following:

$$\lambda K(h) = E\ (\#\ events\ distance\ h\ of\ an\ arbitrary\ event)$$

where λ is called the intensity of the spatial process and is equal to the mean number of events per unit area, a value that is assumed constant over the region of interest. E is the expectation operator, so that the right hand side is the average number of events in the neighborhood of a given event.

When the observed K value is larger than the expected K value for a particular distance, the distribution is more clustered than a random distribution at that distance. When the observed K value is smaller than the expected K value, the distribution is more dispersed than a random distribution at that distance. When the observed K value is larger than the Higher Confidence Envelope (HiConfEnv) value, spatial clustering for that distance is statistically significant. When the observed K value is smaller than the Lower Confidence Envelope (LwConfEnv) value, spatial dispersion for that distance is statistically significant (Figure 30).

Figure 30. Ripley's K-function Spatial Statistics

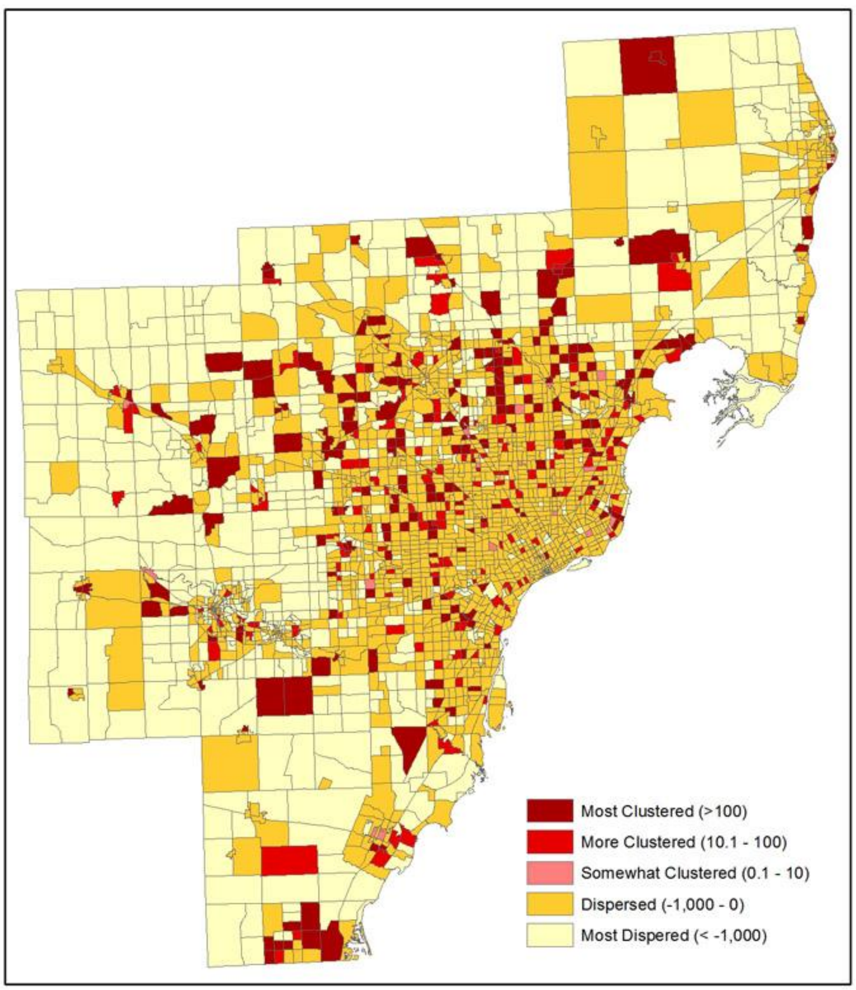


Source: Modified Chart from ArcGIS Manual

The Ripley's k-function statistic is very sensitive to the size of the study area. Identical arrangements of points can exhibit clustering or dispersion depending on the size of the study area enclosing them. Therefore, it is imperative that the study area boundaries are carefully considered. Similarly, the nearest neighbor function is very sensitive to the study area as well. Small changes in the area can result in considerable changes in the results.

The present study tests two sets of boundaries, municipalities and Traffic Analysis Zones (TAZs), to measure the degree of clustering activities. Both ANN scores and Ripley's K-function statistics were calculated within both municipalities and TAZs. For Ripley's K-function statistics, in order to assign a cluster index to each municipality or TAZ, number of distance bands was set to one. Both the difference (Figure 31) and the ratio between observed K value and expected K value were calculated and tested as measures for degree of clustering. I chose the difference measure for the final models because it shows more effectiveness than other cluster measures.

Figure 31. Ripley's K, Difference between Observed and Expected, by TAZ



E. Expected Model Results

As discussed in the Section A of this chapter, this study estimates multinomial logit models for analyzing residential location choice based on household attributes and place characteristics including accessibility. The dependent variable is the probability of residential locations being chosen. For each household in the sample dataset, these locations include the observed chosen location and 29 other randomly sampled locations. The independent variables were selected to describe the factors which influence each household's choice of a location.

1. The Direction of Independent Variables

The direction of the relationship between independent variables and modeled outcomes is given by the sign of the coefficients estimated for the variables. A positive coefficient means that the probability of choosing a location increases as the value of the variable increases, vice versa. T-statistics and adjusted utility (discussed in greater detail later in this section) also carry the same sign. Although the value of the coefficients may be difficult to interpret because variables do not share the same scale or distribution, one may expect local accessibility, mid-range accessibility, and regional accessibility with or without clustering effects should have positive signs, meaning that households are more likely to choose locations with better accessibility.

Some of the control variables are expected to have straightforward effects on the model. The first group of control variables is affordability and economic variables. In this group of variables, first and foremost, “income_-_cost” should be positive because after paying for housing, households should prefer having more financial resources remaining for other consumption needs. For the variables of housing unit size interacting with household income, one would expect that higher income households would be more able to compensate for higher housing prices and afford larger houses. Therefore the coefficients for higher income households should be positive. But it is somewhat uncertain if the coefficients for lower income households can hold positive for this interaction variable (Table 14).

The second group of variables is about social economic characteristics of the households and the neighborhood characteristics. Households tend to live close to similar households. Therefore, interaction variables between household race and percent of same race households in the neighborhood, i.e. Traffic Analysis Zone (TAZ), are expected to be positive. For the same reason, households of various income levels tend to locate with households of similar income levels in general. Interaction variables between households of a certain income level and the percent of same level income households in the neighborhoods are expected to be positive too.

The third group of control variables is about amenities and local services. School quality is expected to have a positive coefficient, especially for high income households. This should be particularly true for households with children. But even for households without children, choosing a neighborhood with good schools can be attractive too for maintaining and increasing housing value and other social qualities.

Table 14. Expected signs of variables

Variables	Signs
1. Affordability/Economic Variables	
income_-_cost	+
housing_size_x_high_inc	+
housing_size_x_mid_inc	?
housing_size_x_low_inc	?
2. Social Composition Variables	
race1_x_zonal_pct_race1	+
race2_x_zonal_pct_race2	+
race3_x_zonal_pct_race3	+
race4_x_zonal_pct_race4	+
high_inc_x_zonal_pct_high_inc	+
low_inc_x_zonal_pct_low_inc	+
3. Amenity and Service Variables:	
school_quality_x_high_inc	+
crime_rate_violent	-
crime_rate_property	-
crime_rate_both	-
crime_rate_both_x_high_inc	-
property_tax	?
is_detroit	?
4. Accessibility Variables	
local_acess_x_young_hh	+
emp_access_x_workers>cars	+
emp_access_x_workers<=cars	+
mid_range_access	+
cluster_k	+
cluster_k_x_emp	?
worker1_commute_time	-
worker2_commete_time	-

Crime is expected to be negative for both violent crimes and property crimes. This should be particularly true for high income households, because they are able to avoid high crime areas with the financial means that they have.

Impacts of property tax on location choice may not be straightforward. While avoiding high tax seems preferable for economic considerations, higher property tax may mean better local services. The expected sign of this variable is thus ambiguous.

The final group of variables is the accessibility variables, which is the focal point of this study. Local accessibility may have different effects on different population groups. Because local accessibility in this study measures accessibility by non-motorized transportation mode such as walking, it may be more attractive to younger households and households without children, who tend to live in walkable urban areas as opposed to households with children who tend to live in lower density suburban areas where houses are larger. Therefore, local accessibility interacting with young households is expected to be positive.

Mid-range accessibility measures daily life attractions within 10 minutes of driving time. It is also expected to be positive because households should prefer more and convenient opportunities for shopping, entertainment, schooling, and other services.

The regional accessibility variables measure employment opportunities. Even though these are not household workers' personal accessibility, households may prefer high employment accessibility in general for such reasons as the ease of finding a job, changing employers, or minimizing total commute for all workers in multi-worker households. Therefore regional accessibility to employment is expected to be positive.

Individual workers' personal commute time may be one of the most important factors for household location choice. One should expect the utility of such variables measuring travel time to work to be negative, because longer commute to jobs makes a residential location less desirable, all else being equal. A significant negative coefficient for workers' commute time will support the notion that workplace accessibility is important in residence location choice.

The indicator of clustering is defined as the difference between the observed Ripley's K value to the expected Ripley's K value by TAZ. It should be positive, because clustering of activities is expected to increase accessibility. The effect of the interaction between clustering and the amount of jobs is unclear. It is reasonable to think that high clustering and a high number of jobs should increase accessibility, therefore a positive sign is expected. On the other hand, areas that show high degree of clustering may consist of a large number of clustered small establishments, whereas areas with large employment numbers may be dominated by a few establishments. In this case, the sign should be negative.

2. The Significance of Variables

The statistical significance of any given variable increases when the absolute value of the t-statistic increases. This helps identify which variables are genuinely related to the outcome being modeled. In this study, the t-statistic identifies the real effects of accessibility measures on the probability of household location choice. Larger absolute t-values indicate stronger evidence of a variable's significance, but any value above the threshold indicates the variable is admissible. Accessibility variables are expected to be statistically significant. UrbanSim, which is the software used for model estimation in this study, provides the suggested t-value. This t-value is based on Bayes Factors. It provides the "weak level evidence" (Raftery 1995) that a variable should be included in a model. An advantage of this form of t-value is that it is directly related to the sample size.

As sample size increases, a higher t-value is needed to show statistical significance of a variable.

3. The Relative Influence of Variables

While it is impossible to directly compare the coefficient of one variable with that of another in the proposed residential location choice model, a series of utility calculations can be performed to help reveal the relative influence of each variable on residential choice. In UrbanSim, a pair of utilities can be calculated using the estimated coefficients for each variable. One utility is determined using the 5th percentile value of that variable and the other using the 95th percentile while all other variables are held constant at the median values. The difference in the two utilities between the 5th and 95th percentile values indicates the influence of that variable. Variables with larger utility value difference contribute more to the result than those with smaller values.

This study expects that the influence of regional accessibility weighted by clustering will be greater than conventional regional accessibility in location choice model. It also expects that adding a mid-range accessibility will increase the combined influence of accessibility (local, regional, and mid-range) on location choice than models without a mid-range accessibility. Furthermore, workers' commute time may have the greatest impact on residential location choice.

4. The Explanatory Power of the Model

The combined explanatory power of a model as a whole can be assessed using the log-likelihood ratio statistics for multinomial logit models. This study expects the model's explanatory power on residential location choice will increase after adding accessibility measures, including mid-range accessibility, incorporating clustering effects into regional accessibility, and individual workers' travel time to work.

F. Summary of Chapter IV

This chapter discussed the methodology of measuring the effects of accessibility on residential location choice. In order to adapt discrete choice theory to analyzing household location choice by buildings on land parcels in this study, a couple of sampling issues are addressed first. Household travel survey needs to be a statistically representative sample. The choice set needs to be of a practical size.

Then the discussion focused on defining accessibility variables as well as control variables for estimating residential location choice models. This was followed by describing the methods of measuring accessibility at various scales and characterizing spatial clusters of activities. Finally the expected modeling results were discussed, particularly for the signs of the coefficients of the control variables and accessibility variables based on theories and intuitions.

CHAPTER V

MODEL RESULTS

This study develops a number of multinomial logit models to analyze the effects of accessibility on residential location choice. The modeling work begins with using socio-economic control variables only as independent variables, then adding various accessibility measures. The results of these models are presented first in this chapter.

Additional models are developed and presented later in this chapter. In order to explore the effects of commute time on residential location choice at various geographic scales, three sets of multinomial logit models are developed at regional, county, and city levels respectively. Furthermore, since Detroit region is one of the most racially segregated regions in the United States in terms of Black and White populations, specific residential location choice models are estimated for Black and White populations respectively in order to assess the differences in residential location choice, explore possible reasons for explaining the differences, and help developing policies.

A. Initial Model Results

A series of multinomial logit models are estimated using the households from 2004-05 SEMCOG household travel survey to explore the effects of various aspects of accessibility on residential location choice. The results of these models are summarized in Table 15. The first model (labeled as Model 1) is a base model that has only control variables as its independent variables, without any accessibility variables. Model 2 has local and regional accessibility measures as independent variables in addition to control variables. Model 3 has mid-range accessibility as an additional independent variable besides all the variables in Model 2. Model 4 incorporates clustering effects. And finally Model 5 has workers' commute time as additional independent variables.

The estimated coefficients for the control variables in Model 1 through Model 5 are consistent with theoretical and intuitive expectations. Beginning with economic and housing affordability variables, this model shows that the estimated coefficients for the interaction variable between household income and housing price ("income_-_cost") to be consistently positive and statistically significant, which is theoretically sound. The variables of income and housing unit size interactions are also consistent with expectations. The signs of the estimated coefficients are positive for all three income groups, meaning that all households view housing size as a positive utility. The magnitude of the coefficients for the three income groups increases as income increases, indicating that the utility of housing size is higher for higher income households, whereas the utility of housing size is lower for lower income households due to financial constraints. It reflects the economic ability for households of various income levels to pay for larger housing.

Table 15. Results of Residential Location Choice Models

	Model 1		Model 2		Model 3		Model 4		Model 5	
Variables	**Coeff.**	**t-value**	**Coeff.**	**t-value**	**Coeff.**	**t-value**	**Coeff.**	**t-value**	**Coeff.**	**t-value**
1. Affordability/Economic Variables										
income_-_cost	0.0718	10.25*	0.0748	10.54*	0.0699	10.06*	0.0734	10.40*	0.0653	8.59*
housing_size_x_high_inc	0.7246	8.90*	0.6991	8.39*	0.7183	8.74*	0.7408	8.88*	0.7427	8.33*
housing_size_x_mid_inc	0.3607	8.34*	0.3691	8.55*	0.3530	8.22*	0.3815	8.82*	0.3658	8.19*
housing_size_x_low_inc	0.3172	5.82*	0.2943	5.37*	0.2712	4.98*	0.2847	5.22*	0.1891	3.34*
2. Social Composition Variables										
race1_x_zonal_pct_race1	0.0230	17.04*	0.0245	18.12*	0.0239	17.55*	0.0242	17.31*	0.0242	16.57*
race2_x_zonal_pct_race2	0.0345	23.97*	0.0332	23.08*	0.0348	23.89*	0.0346	23.97*	0.0347	23.17*
race3_x_zonal_pct_race3	0.0690	12.00*	0.0606	10.88*	0.0564	9.72*	0.0689	11.76*	0.0724	10.86*
race4_x_zonal_pct_race4	0.0912	10.54*	0.0860	9.98*	0.0866	10.69*	0.0894	10.05*	0.0992	9.61*
high_inc_x_zonal_pct_high_inc	0.0147	8.13*	0.0154	8.38*	0.0155	8.39*	0.0159	8.55*	0.0150	7.41*
low_inc_x_zonal_pct_low_inc	0.0339	12.93*	0.0319	12.01*	0.0319	12.15*	0.0315	11.92*	0.0312	11.44*
3. Amenity and Service Variables:										
school_quality_x_high_inc	0.0038	0.98	0.0050	1.31	0.0032	0.82	0.0028	0.73	0.0042	0.98
property_tax	0.0299	12.18*	0.0239	9.09*	0.0184	6.92*	0.0171	6.22*	0.0201	6.55*
is_detroit	-0.8046	-8.58*	-0.6940	-7.36*	-0.6101	-6.50*	-0.5837	-6.09*	-0.6642	-6.54*
4. Accessibility Variables										
local_access_x_young_hh			0.1278	2.84	0.1284	2.81	0.0823	1.86	0.1135	2.34
emp_access_x_workers>cars			9.9E-06	4.28*	7.9E-06	3.48*	9.5E-06	4.27*	1.5E-06	0.61
emp_access_x_workers<=cars			7.2E-07	5.83*	3.4E-07	2.72	2.2E-07	1.35	-3.6E-07	-1.20
mid_range_access					0.1244	9.30*	0.1171	8.56*	0.0607	4.00*
cluster_k							1.5E-07	1.02	2.5E-07	1.35
cluster_k_x_emp							-8.6E-10	-1.95	2.0E-07	3.70*
worker1_commute_time									-0.0945	-46.84*
worker2_commute_time									-0.0777	-34.46*
Log-Likelihood	-12477.3		-12468.1		-12427.9		-12391.8		-9880.0	
Null Log-Likelihood	-14366.7		-14366.7		-14366.7		-14366.7		-14366.7	
Log-Likelihood Ratio (ρ)	0.13151		0.13215		0.13495		0.13746		0.31230	

Notes: (1) Number of observations is 4,224. (2) Suggested |t-value| is 2.89. (3) “*” indicates statistically significant coefficients.

The next group of variables is about demographic and socio-economic characteristics of the households and their neighborhoods. The coefficients for the race variables are all positive, indicating that for all races, households are more likely to choose neighborhoods (represented by Traffic Analysis Zones for these four variables) where there are more households of the same race. This is an indication of people's preference to locate close to same race neighbors as well as other factors for racial segregation, which will be further discussed later in this chapter. The two income interaction variables show similar results as the race variables, i.e., high income households tend to locate close to other high income households, whereas low income households tend to locate close to other low income households.

The third group of control variables is about local services and amenities. School quality has a positive utility for high income households. They prefer to locate in good school districts with the financial resources that they have. However, the present models do not show this variable being statistically significant. This is in part due to the high correlation between school quality and other income related variables in the model. Property tax is positive in the present models, indicating its functioning as a proxy to services provided by the local governments that have positive effects on the utility functions. The Detroit dummy variable is used to capture the factors that cannot be observed in the central city. It shows a negative utility for household location choice.

Despite numerous attempts to include several forms of crime variables in the models, they failed to provide meaningful additional explanation of residential location choice. The main reason is that crime variables are highly correlated (negatively) to school quality variables. And they are both correlated to all the other income-related variables. School quality variable performed more consistently in the models than crime rates. Therefore, school quality variable is included in the models whereas crime rate variables are left out of the present models.

Overall, the estimated coefficients for the control variables in Model 1 through Model 5 confirmed their expected effects on residential location choice. All the coefficients are statistically significant except for one that is the interaction variable of school quality and household income. Nevertheless, it at least has the right sign, showing the positive impact of school quality to high-income households' location choice.

Model 2 adds local and regional accessibility measures in addition to all the control variables in Model 1. Local accessibility is represented by the composite index as described earlier in the last chapter. This is a robust measure of local accessibility which incorporates the density, diversity, and design all three aspects of local accessibility. The variable itself for all households showed little impact on residential location choice, and was excluded from the model. However, the interaction between this measure and the young households shows a positive impact on location choice. And the t-value for the coefficient is almost statistically significant. This confirms that young people are more likely to locate in areas that have high local accessibility.

Regional accessibility to employment is split into two variables. They both are based on logsums. One of them measures regional accessibility for households that have fewer motor vehicles available than the number of workers in the households. The other one measures regional accessibility for households that have at least the same number of vehicles available as number of workers in the households. Both variables have statistically significant positive coefficients, confirming that regional accessibility to employment is significant for residential location choice.

Model 3 adds mid-range accessibility in addition to all the variables in Model 2. The mid-range accessibility variable is a cumulative opportunity measure that counts number of retail, leisure, education, and other services jobs within a household's 10 minutes driving radius. The positive coefficient of this variable indicates that accessibility to nonwork destinations at sub-regional market areas has a positive impact on household location choice. This variable is statistically significant. After this variable is added into the model, the coefficient for local accessibility interacting with young households increases slightly, and its t-value decreases slightly but remains close to be statistically significant. The coefficients and t-values for regional accessibility variables decrease. Note that for households that have more workers than motor vehicles available, the coefficient for their regional accessibility to employment is more than 10 times higher than the same coefficient for households with at least same number of vehicles as number of workers, and its t-value remains statistically significant. This indicates that regional accessibility is more important to households with insufficient motor vehicles available to their workers and other members in the households than households with sufficient vehicles available. The results of Model 3 essentially confirms the first hypothesis of this study, that is a sub-regional mid-range accessibility has a statistically significant effect on residential location choice, controlling for regional and local accessibility.

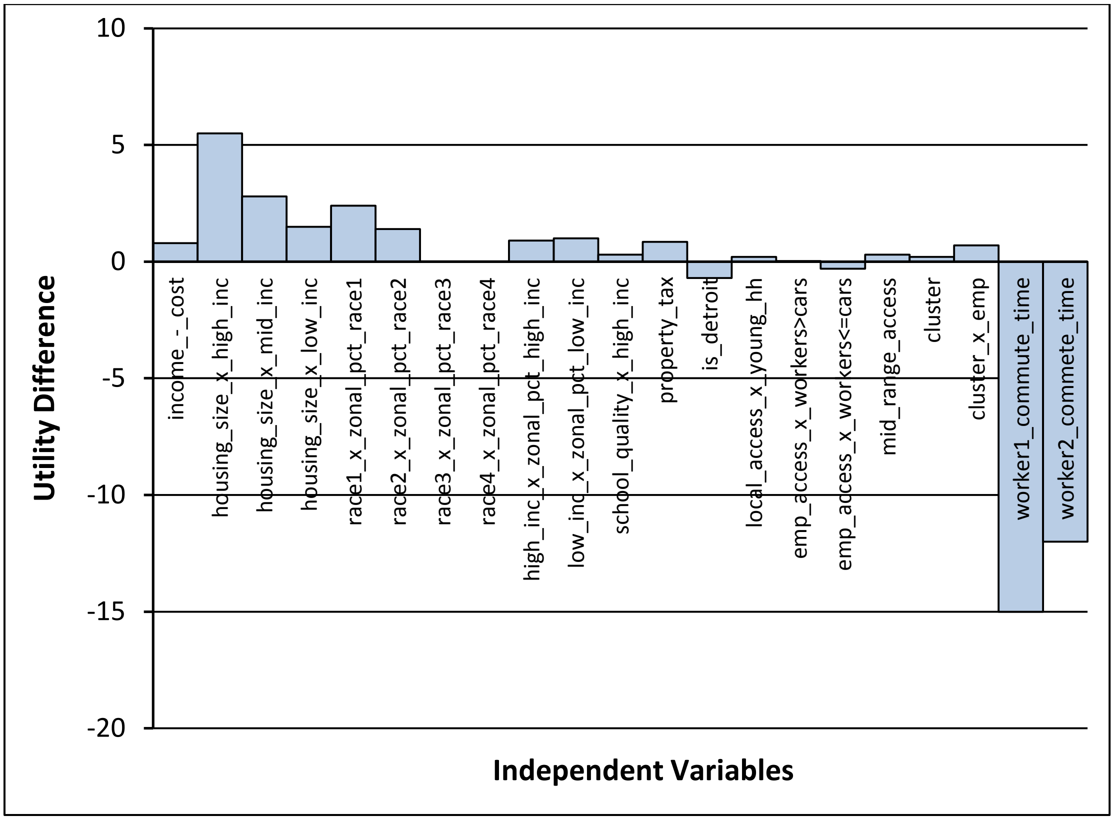
Model 4 incorporates clustering effects of destinations. Besides the amount of employment by TAZ, cluster scores by TAZ (measured by the observed Ripley's K statistics minus the expected K statistics), and the interaction between the amount of employment and cluster score for each TAZ are included in the model, to assess the effects of regional accessibility on residential location choice from all these three related variables. The coefficient for the amount of employment remains significant (t-value = 4.27) for households with fewer vehicles than workers. The coefficient for the clustering itself has a positive sign, indicating that clustering has a positive effect on location choice. But it is not statistically significant (t-value = 1.02). The interaction between the amount of employment and clustering has a negative sign, which could be caused by the fact that smaller establishments tend to be more clustered than larger establishments. But it is also not statistically significant (t-value = -1.95). The model's explanation power increased somewhat indicated by the increase of the model goodness of fit measure, log-likelihood ratio (ρ). The expected effect as stated in the second hypothesis is that models incorporating clustering will show stronger explanatory power in predicting residential location choice than models that do not incorporate clustering. While the model result shows some increase in explanation power, and the sign for clustering is right, the coefficients for clustering and for the interaction between clustering and amount of employment are far from being statistically significant. Therefore, the result from the modeling work for the second hypothesis shows some encouraging signs, but it is inconclusive in the present study. Further research is needed in the future.

Model 5 adds individual workers' commute time as an additional accessibility measure. This model includes travel time to work for up to two workers in each household. The estimated coefficients for both workers are negative and highly significant statistically. It indicates that workers' travel time to work is a negative utility for household location choice. Households prefer home locations that are closer to jobs.

Mid-range accessibility remains significant in Model 5, and local accessibility for young households is close to being significant. Clustering and clustering interacting with employment amount are both positive. And the coefficient for the interaction variable is statistically significant. The drawback of adding commute time in the current form of Model 5 is that place-based regional accessibility to the amount of employment is no longer significant. Logsum-based accessibility to employment for households with more workers than available vehicles remains positive but not significant. The impact of commute time is so large that the sign is changed for logsum-based accessibility to employment for households with same or more vehicles available than workers. However, the most important finding in this modeling work is that adding workers' commute time more than doubled the model's log-likelihood ratio, which is the goodness-of-fit measurement. This means that commute time variables significantly increase the model's explanation power. Overall, the modeling result confirms the third hypothesis of this study in the sense that individual workers' commute time is highly significant in residential location choice and dramatically increases the explanation power of the models. However, some of the place-based accessibility measures cannot be found significant simultaneously.

Although it is impossible to directly compare the coefficient of one variable with that of another in these multinomial logit models, a series of utility calculations can help reveal the relative influence of each variable on residential location choice, as explained in the last chapter. The relative influence of each variable in Model 5 that is calculated by such utility difference is shown in Figure 32.

Figure 32. Relative Influence of Independent Variables



In Figure 32, commute time for worker 1 and worker 2 have the greatest influence on residential location choice. Their influence is much greater than any other variables. This suggests that accessibility to jobs is the most important factor for residential location choice among all factors that have been identified and measured in this study. It also suggests that commute time is the most effective accessibility measure in modeling residential location choice.

B. Regional, County, and City Models

One may wonder how geographic scales may affect the significance of accessibility on residential location choice. It might be easier to imagine that travel time to work matters in residential location choice at the regional scale. Regions are large. They may extend over 100 miles from one side to the other. People normally do not travel that far to go to work. People tend to live relatively closer to jobs when the entire region is considered to be the choice set. What happens if only a portion of the region is considered in a discrete choice model? Will the effects of accessibility on residential location choice decrease when most of the jobs are within reasonable commuting time? Will the journey-to-work effects on location choice completely disappear if the study area becomes very small, and all that matters is the other factors such as housing characteristics? To answer these questions, this study estimates a set of residential location choice models at three different geographic levels in the Detroit region. They are the entire Detroit region, Washtenaw County, and City of Ann Arbor respectively. They vary greatly in terms of area, number of households, population, and jobs (Table 16).

Table 16. Three Levels of Geographic Scales

Study Areas	Land Area (Acres)	Households	Population	Jobs
Detroit Region	2,946,745	1,844,758	4,704,743	2,484,251
Washtenaw County	462,248	137,193	344,791	236,676
City of Ann Arbor	17,986	47,060	113,934	120,588

Data Source: SEMCOG

There are seven counties in the Detroit region. Washtenaw County is selected not only because its size in terms of population and jobs is in the middle of the seven counties, but more importantly because it functions like an independent small region by itself with a core city, suburban neighborhoods, and rural townships. The City of Ann

Arbor is the central city of Washtenaw County with abundant jobs and housing opportunities to choose from, and it has a good size of household samples from the 2004-05 SEMCOG travel survey for model estimation.

For the Washtenaw County model and Ann Arbor City model, only households that live and work in these areas respectively are considered. These models are about residential location choice. They obviously should not include households that do not live in the study area. Furthermore, the purpose of estimating these models is to assess the difference among models at various geographic scales. If jobs outside the study areas were included, the model would still be a regional model for a subset of choosers who happened to live in Washtenaw County or City of Ann Arbor. Therefore, households must live and work in the study area in order to study the scale effects on the models.

To ensure consistency among the models, the same set of variables is used in all the models (Table 17). There is a pair of models for each one of the three geographic levels. Each pair has a model with control variables only and a model with control variables and workers' travel time to work, i.e., commute time.

The modeling results clearly show that travel time to work is significant at all the three geographic levels. It might seem to be surprising that the absolute value of the coefficients of travel time to work increases when the size of study areas decreases. In fact, this indicates that commuters are more sensitive to travel time when areas become smaller. There is a strong intuitive reason for people to be more sensitive to changes in travel times when the size of the area in consideration decreases.

Table 17. Model Results at Three Geographic Levels

Variables	Model 6: Detroit Region				Model 7: Washtenaw County				Model 8: Ann Arbor City			
	Without Travel Time		With Travel Time		Without Travel Time		With Travel Time		Without Travel Time		With Travel Time	
	Coeff.	t-val.	Coeff.	t-val.	Coeff.	t-val.	Coeff.	t-val.	Coeff.	t-val.	Coeff.	t-val.
Control Variables												
income_-_cost	0.0761	13.07*	0.0752	12.00*	0.1061	4.83*	0.1257	5.55*	0.1596	4.62*	0.1630	4.44*
housing_size_x_high_inc	0.6591	11.67*	0.7325	12.42*	0.5828	2.37	0.5887	2.36	0.6076	1.95	0.6547	2.02
housing_size_x_mid_inc	0.3813	10.75*	0.4420	11.87*	0.2927	1.46	0.4915	2.48*	0.6690	2.01	0.6469	1.92
housing_size_x_low_inc	0.2624	5.81*	0.2209	4.75*	0.2466	1.52	0.3910	2.34	0.3698	0.93	0.4150	1.11
race1_x_zonal_pct_race1	0.0296	27.92*	0.0346	30.51*	0.0244	5.62*	0.0320	7.20*	0.0120	1.49	0.0057	0.68
race2_x_zonal_pct_race2	0.0319	31.59*	0.0281	25.86*	0.0331	4.15*	0.0358	4.35*	0.1224	3.93*	0.1533	4.10*
race3_x_zonal_pct_race3	0.0618	11.59*	0.0588	9.69*	0.2434	3.75*	0.2526	1.99	0.4430	1.60	0.3474	2.08
race4_x_zonal_pct_race4	0.0918	11.90*	0.0862	9.71*	0.0642	3.58*	0.0486	2.73*	0.0115	0.52	0.0154	0.79
high_inc_x_zonal_pct_high_inc	0.0141	10.01*	0.0150	9.79*	0.0209	4.28*	0.0179	3.66*	0.0265	3.13*	0.0278	3.07*
low_inc_x_zonal_pct_low_inc	0.0373	17.49*	0.0362	15.92*	0.0403	6.80*	0.0381	6.54*	0.0287	3.75*	0.0236	2.87*
Accessibility Variables												
worker1_commute_time			-0.0820	-58.4*			-0.0860	-10.8*			-0.1415	-3.7*
worker2_commete_time			-0.0679	-44.3*			-0.0683	-7.04*			-0.1184	-2.5*
Number of Observations	5682		5682		460		460		177		177	
Log-Likelihood	-16957.3		-13404.5		-1447.9		-1365.7		-555.7		-543.7	
Null Log-Likelihood	-19325.6		-19325.6		-1564.6		-1564.6		-602.0		-602.0	
Log-Likelihood Ratio (ρ)	0.12255		0.30639		0.07455		0.12710		0.07687		0.09694	
Adjusted ρ (ρ')	0.12203		0.30577		0.06815		0.11943		0.06026		0.07700	
Change of ρ			0.18384 (150.0%)				0.05256 (70.5%)				0.02006 (26.1%)	
Change of ρ'			0.18374 (150.6%)				0.05128 (75.2%)				0.01674 (27.8%)	
Suggested |t-value| >	2.94		2.94		2.48		2.48		2.28		2.28	

Note: Statistically significant variables are indicated by "*".

At the regional scale, adding a minute of commuting time may seem trivial, whereas adding a minute of commuting time in the City of Ann Arbor may result in greater marginal disutility compared to the regional level. This argument can be further verified by the so called "friction factors" in transportation models. In the gravity model based trip distribution models, friction factors indicate the level of impedance to interact, or the unwillingness to travel. The friction factor increases for home-based work trips when study area decreases (Table 18), which indicates greater impedance in smaller areas and confirms that workers are more sensitive to travel time with greater unwillingness to take additional time to commute.

Table 18. The Friction Factors of Three Levels of Geography

	Region	Washtenaw County	Ann Arbor
Home-based Work Trips	0.111	0.126	0.175
Home-based Shopping Trips	0.301	0.298	0.388
Home-based School Trips	0.387	0.398	0.377
Home-based Other Trips	0.240	0.245	0.315

One may argue that people who live and work in Ann Arbor are a "self-selected" group who are less willing to commute compared to other commuters. This may in part be true. But it does not void the basic argument of this discussion that travel time to work is a significant factor in residential location decisions.

The changes of the model goodness of fit measure, log-likelihood ratio (ρ), in Table 17 provides additional insight. The ratio decreases when study area decreases, indicating greater unobserved randomness in smaller areas. Furthermore, the additional explanatory power of travel time to work decreases when study area decreases as indicated by the percent change of log-likelihood ratio (ρ) within each pair of models that decreases when study area decreases. For the regional model, adding commute time to the control model increases log-likelihood ratio by 150%. The percent change decreases to 70% and 26% for Washtenaw County model and Ann Arbor City model respectively. Nevertheless, these results indicate that workers' travel time to work affects residential location choice at various geographic scales, from a large region, to a mid-sized county, to a specific municipality. To the knowledge of this author, this is the first study that demonstrates such effects at multiple geographic scales.

C. Model by Race

The model estimations discussed so far in this study use population attributes in the models as independent interaction variables. Another way to analyze the effects of various population groups' accessibility on location choice is to estimate a separate model for each group, such as by race and ethnicity. Racial discrimination may limit people's ability to optimize freely their residential locations. This is particularly relevant in regions like Detroit where racial segregation is high. Accessibility varies among population groups. Although Black-White segregation is declining fairly consistently for most metropolitan areas, many changes are small, preserving the long-standing segregation for different racial and ethnic groups (Frey and Myers 2005). The level and change of segregation can be attributed to a variety of demographic and economic contextual factors in a region. These factors include disparate economic resources across groups, preferences to reside with neighbors of the same group, municipal zoning laws that discourage racial and economic integration, and the long history of discriminatory practice on the part of real estate agents, rental agents, lending institutions, and insurers.

A series of trend studies (Taeuber and Taeuber 1965; Massey and Denton 1993) have documented that the effects of discriminatory practice is most evident in the segregation of African Americans from Whites. Frey and Farley (1996) found that, on average at the national level, six out of ten Blacks would have had to change neighborhoods (by census block groups) in order to be distributed in the same way that Whites were. Meanwhile, Hispanics and Asians were also segregated but less substantially than Blacks. On average, four in ten Hispanics or Asians would have had to change residence to be distributed like the White population in their respective metropolitan areas.

The Detroit region is one of the regions in the United States that have a large number of African Americans and have highly segregated Blacks and Whites residential location pattern. In fact, the Detroit region ranks the second in Black-White segregation (Frey and Myers 2005). It is desirable to model location choice for Blacks and Whites in this region separately, compare the differences, search for reasons, and help making policies to reduce segregation. This study estimates two alternative residential location choice models based on Model 5, one for Whites and the other one for Blacks. The modeling results are summarized in Table 19.

Table 19. Residential Location Choice Models for Whites and Blacks

	Model 9: Whites		Model 10: Blacks	
Variables	**Coeff.**	**t-value**	**Coeff.**	**t-value**
1. Affordability/Economic Variables				
income_-_cost	0.0788	8.82*	0.0514	3.30*
housing_size_x_high_inc	0.8314	8.37*	0.3446	1.25
housing_size_x_mid_inc	0.3578	6.79*	0.3907	4.35*
housing_size_x_low_inc	0.1609	2.46	0.3684	3.08*
2. Social Composition Variables				
race1_x_zonal_pct_race1	0.0218	12.84*	na	na
race2_x_zonal_pct_race2	na	na	0.0334	16.36*
high_inc_x_zonal_pct_high_inc	0.0106	4.85*	0.0192	2.77*
low_inc_x_zonal_pct_low_inc	0.0272	9.25*	0.0275	5.95*
3. Amenity and Service Variables:				
school_quality_x_high_inc	0.0061	1.25	-0.0206	-1.95
is_detroit	-0.4983	-3.38*	-0.1622	-1.18
4. Accessibility Variables				
local_access_x_young_hh	0.3263	5.54*	-0.1201	-1.13
mid_range_access	0.0767	4.64*	0.0590	1.21
emp_access_x_workers>cars	5.40E-06	0.61	1.04E-05	0.76
emp_access_x_workers<=cars	-2.21E-06	-9.18*	1.11E-07	0.19
worker1_commute_time	-0.0913	-41.70*	-0.1104	-16.46*
worker2_commute_time	-0.0757	-31.21*	-0.0933	-11.29*
Number of Observations	2,940		996	
Log-Likelihood	-7180.37		-2063.06	
Null Log-Likelihood	-9999.52		-3387.59	
Log-Likelihood Ratio	0.28193		0.39099	
Adjusted Log-Likelihood Ratio	0.28043		0.38657	
Suggested |t-value| >	2.83		2.63	

Note: Statistically significant variables are indicated by "*".

Economic disparity is one major reason for racial segregation. Coefficient differences of a few variables in Model 9 and 10 seem to reflect the impact of economic disparity. Note that the most influential variables of workers' commute time, coefficients for Blacks are more negative than for Whites for both workers in a household. This may be an indication that Blacks are less able to afford transportation costs than their more affluent White counterparts. Similar logic may apply to the place-based regional accessibility measures. Coefficients for these variables are more positive for Blacks than for Whites. It shows Blacks are more favorable to high accessibility areas than Whites, which could be again an effect of their inability to pay higher transportation costs. By contrast, White young households have a statistically significant positive coefficient for local accessibility whereas Black young households have a negative coefficient for that variable. Areas with high local accessibility where young households like to live tend to be more expensive places. The premium costs may be prohibiting poorer young Black households to live in those places.

Economic disparity cannot explain all the differences in coefficients for Blacks and Whites. For example, school quality has a positive coefficient for Whites but a negative coefficient for Blacks, even though those are all high income households. Explanations may come from preference to reside with same race neighbors, and/or institutional discriminatory practices that may force Black households to live in areas with lower school qualities. These same forces may explain the coefficient difference for same race variables. The coefficient for Blacks to live in higher percentage Blacks neighborhoods is 0.0334 (t-value = 16.36), which is higher than the coefficient for Whites to live in higher percentage Whites neighborhoods (coefficient = 0.0218, t-value = 12.84).

D. Summary of Chapter V

This chapter discussed the modeling results which clearly show the important effects of accessibility on residential location choice. The three research hypotheses of this dissertation are supported to various degrees. First, the sub-regional mid-range accessibility has a statistically significant effect on residential location choice, controlling for regional and local accessibility. Second, models that incorporate clustering effects show somewhat stronger explanatory power in predicting residential location choice than models that do not incorporate clustering effects. However, the improvement is small. Third, individual workers' commute time is highly significant in residential location choice and is the most powerful variable in the model, although some place-based accessibility measures cannot be found significant simultaneously. Adding commute time to residential location choice model more than doubled the explanatory power of the model.

The significant effects of commute time on residential location choice were found at regional, county, and city levels. To the knowledge of this author, this is the first study that demonstrated the significance of commute time on residential location choice at multiple geographic scales within a region. Additional modeling results show that the effects of accessibility on residential location choice vary among population groups such as by race and ethnicity. The next chapter will discuss the implications of this study from three perspectives: planning theory, modeling practice, and public policy development.

CHAPTER VI

DISCUSSIONS AND CONCLUSIONS

The final chapter of this dissertation provides discussions on the implications of the modeling results, suggestions for future research areas, and conclusions of the dissertation.

The implications of the modeling results are discussed in three perspectives: theoretical clarifications, practical improvements, and public policy development. From the theoretical point of view, the fact that workers' commute time has consistently shown statistical significance suggests the important effects of accessibility on residential location choice. In planning practice, the results of this study should help improve the accuracy of land use and transportation models. This will be demonstrated by running the estimated household location choice models to predict home locations of the households from the SEMCOG travel survey, and comparing the predicted home locations to their observed locations. The results of this study support public policies that take into account the important relationship between land use and transportation interactions.

After discussing the implications of the modeling results, this chapter then suggests future research areas including additional data and improvement of methods. The chapter completes with conclusions for the dissertation by highlighting the findings of this study.

A. Theoretical Implications

The issue of accessibility's effects on residential location choice is controversial. There has been contradictory empirical evidence to support or dismiss such effects. Studies based on excess commuting approach divides "actual commuting time" into "required minimum commuting time" and "excess commuting time." By showing that the amount of "required minimum commuting time" is less than the amount of "actual commuting time," Giuliano and Small (1993, p. 1485) stated that "other factors must be more important to location decisions than commuting cost, and that policies aimed at changing the jobs-housing balance will have only a minor effect on commuting."

However, this study finds that accessibility significantly affects residential location choice. Particularly, workers' travel time to work is found to be *the most* influential factor in choosing a residential location. This funding is based on comprehensive multivariate statistical analyses. By contrast, the excess commuting approach uses an unusual bivariate construct in its methodology.

The excess commuting theory is based on the assumption that individual households, each minimizing its commuting cost, will achieve an equilibrium with no "cross-commuting," which is one that minimizes aggregate commuting cost given the distributions of housing and job locations. This cost-minimizing assignment approach is solely based on job locations and resident locations. This is the fundamental problem with the excess commuting approach because it is based on an extremely strong and narrow assumption that dramatically simplifies reality. It is a bivariate analysis that expects commuting itself to fully explain residential location choice decisions. It ignores many other important factors that affect commuting and residential location choice. The null hypothesis is equivalent to an r-squared of "1.0" that commuting time should explain all the variance in residential location choice. Any outcome that is less than "1.0" would result in the conclusion of "wasteful commuting." However, in a typical social science approach, the null hypothesis is typically equivalent to an r-squared of "0.0", meaning no effects of a variable. The analytical task is to prove that the effect of a variable is significantly different from "0.0". This dissertation study takes this more conventional and comprehensive approach and shows that commute time is highly significant in residential location choice, controlling for all the other variables that can be measured by this study. In fact, commuting is the most powerful variable found in this study.

Excess commuting studies produced a wide range of estimates of wasteful commuting, from 11.1% (White 1988) to 87.1% (Hamilton 1982). The size variation of analysis unit is one of the main reasons for the very divergent estimates of excess commuting. The linear programming model uses the origin and destination matrix, which is usually based on Traffic Analysis Zones (TAZs). The size and the number of TAZs used in linear programming models affect the estimation of excess commuting. The high degree of aggregation of zones tends to diminish the proportion of excess commuting because the transportation optimization model does not change jobs or residential places to minimize work trips *within* a zone. In the most extreme case, excess commuting is zero when the number of zones is one for the entire study area, because there are no jobs or households for the linear programming to "swap." Small and Song (1992) further investigated the geographic scale issue in excess commuting analysis. They argued that this aggregation bias is more serious than expected. In their study there was a huge difference of the estimated excess commuting between studies using aggregated and disaggregated data. They found that if large zones are used, only one-third of the actual commuting could be classified as excessive commuting. On the contrary, if small zones are used, about two-thirds of actual commuting could be classified as excessive. This again indicates that the amount of excess commuting is very sensitive to the level of aggregation.

This dissertation also deals with the scale issue, from a different perspective. Multinomial logit models are estimated at three geographic scales: region, county, and city. One might think that commute time could affect residential location choice at the regional scale but not in a smaller geography. However, the result of this study suggests that commute time is the most influential factor on residential location choice at all three geographic scales.

B. Practical Implications

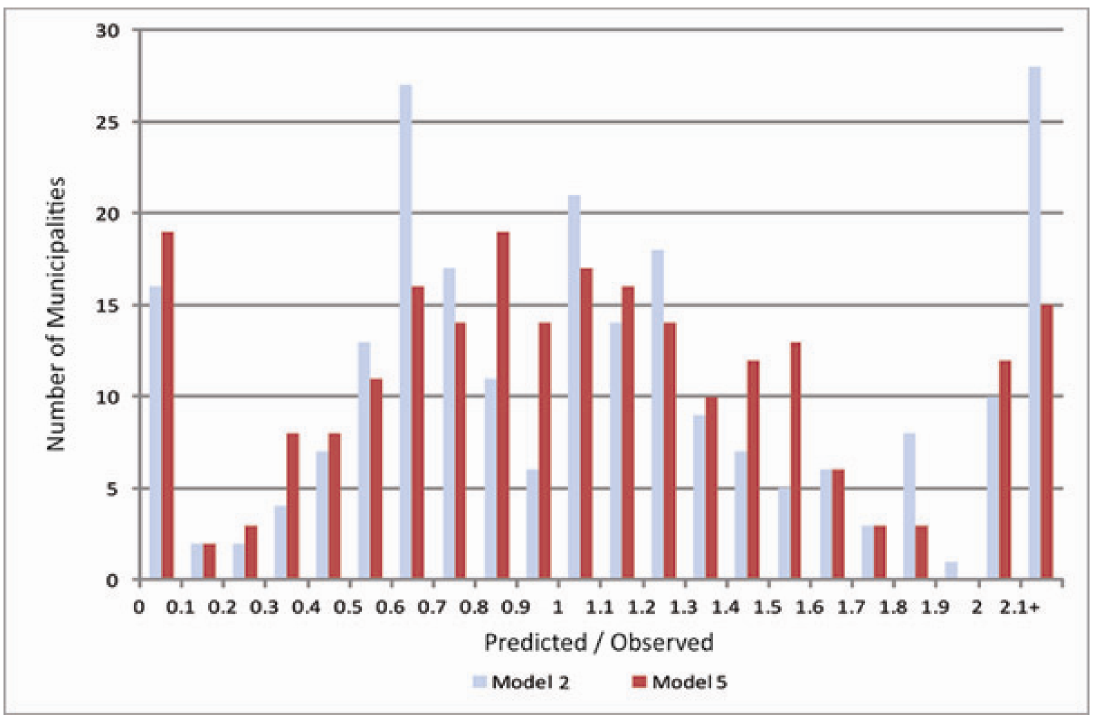
The most direct implication of this study in urban and regional planning practice is the potential to improve land use and transportation modeling, which is a key function of metropolitan planning organizations. Improving the quality of modeling work would provide more accurate information for land use and transportation decision making. Hansen (1959) has been cited often in accessibility research for his seminal work in developing the concept and constructing the basic measures of accessibility. It has been less frequently noted that Hansen’s intention was to develop a residential land use model that relates the accessibility of an area to the rate and intensity of land development in that area “based on a realistic measurement of accessibility.” Exploring accessibility measures from various perspectives and their impacts on residential location choice is the key purpose of this study. A goal of this study is to contribute to improving land use and transportation modeling by studying accessibility at various scales including a mid-range accessibility measure, clustering of destinations, as well as both place-based accessibility and individual workers’ commute time. To demonstrate the technical implications of the findings of this study, I run the base model (Model 2) and the improved model (Model 5) to predict home locations of the households from the 2004-05 SEMCOG household travel survey and compare the predicted home locations to the observed home locations. The comparison is not at specific home location level, because it is almost impossible for any model to pick exactly the same house from over 1.9 million houses in a region for any households. And the multinomial logit model is not a deterministic model, but a probabilistic model. What really matters is to locate the right type of households to the right type of locations. Therefore the comparison is made at the municipality level for this study.

There are 235 municipalities in the Detroit region. Using the predicted number of households to the observed number of households ratio as a measure for assessing prediction success, running Model 2 resulted in 27 municipalities where the prediction error is less than 10% (Table 20). Running Model 5 improved the results to 31 municipalities, an increase of 14.8%. Similarly, running Model 2 resulted in 52 municipalities with a less than 20% error, and running Model 5 improved the result to 66 municipalities. The 26.9% increase is a larger improvement than municipalities with a less than 10% error. A graphic comparison of the complete results from the two models by ten percentage point interval is shown in Figure 33.

Table 20. Number of Households, Predicted/Observed, by Municipality

	Number of Municipalities		Percent of Municipalities		Percent Change
	Model 1	Model 5	Model 1	Model 5	
Within 10% Error (0.9-1.1)	27	31	11.5%	13.2%	14.8%
Within 20% Error (0.8-1.2)	52	66	22.1%	28.1%	26.9%
Within 30% Error (0.7-1.3)	87	94	37.0%	40.0%	8.0%
More than 30%	148	141	63.0%	60.0%	-4.7%

Figure 33. Number of Households, Predicted/Observed, by Municipality



In regard to the households that are exactly predicted to the same municipalities as observed, the results are improved from 22% (Model 2) to 26% (Model 5), which is an improvement of 16.2%. These results confirm that a higher percentage of households can be predicted in the right municipalities by using the improved model. Given the fact that workers' commute times more than doubled the model goodness of fit statistics, these model improvements are expected.

C. Policy Implications

The practical implications of this study indicate the potential of improving land use and transportation modeling to provide better information for policy making. If mid-range accessibility, clustering of destinations, and commute time are relevant, they ought to improve the predictive capacity of land-use and transportation models. Modeling is central to regional land use and transportation planning. Getting a better answer to questions of land-use impacts on transportation and vice versa would give policy makers clearer information for decision making. Better information should lead to improvement of decision making.

The theoretical implications of this study confirm the importance of accessibility in land use and transportation connection. This is contrary to the conclusions drawn in the excess commuting studies which argued that the metropolitan-wide structure of urban land use via policy intervention is likely to have disappointing impacts on commuting patterns. To provide policy guidance, any policy-related interpretations of empirical studies are inevitably dependent on the reliable methods. Policy conclusions drawn from the excess commuting analysis are biased due to its overly simplified view on the complex real world and expecting commuting time to explain everything in location decisions while unable to include other variables in the analysis. Acknowledging the inability to explain why “journey to work play only limited role in residential location choice,” some researchers offered a few “hypotheses” including “job turnover,” “two-worker households,” “non-work trips,” “other priorities” than transportation costs, and “racial discrimination” (Giuliano and Small 1993, p. 1498). But the results of this dissertation research show contradictory evidence to several of these “hypotheses".

First, among all the arguments that commuting time plays a limited and diminishing role in urban policy, one of the most frequently mentioned reasons has been the increasing number of two-worker households, that makes optimizing one worker’s commute may worsen that of the other. While it is intuitive to think that two-worker households may have more constraints than single-worker households, this study has shown that both workers’ travel time to work are statistically significant in residential location choice decisions. Policies that encourage better work accessibility may actually have great potential to improve both workers’ commute.

Second, it has also been widely mentioned that the increasing importance of nonwork trips reduces the influence of work trips on residential location choice. This study first finds that nonwork accessibility oriented mid-range accessibility is significant in residential location choice. But work-based regional accessibility remains significant after adding mid-range accessibility. Furthermore, workers’ travel time to work has much more explanatory power than mid-range accessibility and any other variables that can be measured in this study. Public policies need to address both work and nonwork accessibility.

Third, it has been argued that transportation costs are overshadowed in importance by other factors such as neighborhood characteristics and housing varieties that urban residents care most. This study uses the most detailed and disaggregated available data and concludes otherwise. Housing cost and size are found significant but less powerful than commute time. School quality variable has the right sign in the models but not statistically significant. This is not to say that school quality is not important. But it is highly correlated to income. Since affordability is a more direct and influential variable, it is more significant than school quality when both are in the model. Similar results are found for crime rate variables. They cannot coexist with income and school quality in the models, although they show significant disutility by themselves. On the other hand, workers' commute time explains more variance than any of the amenity variables. Policies that emphasize amenity and services but ignore accessibility seem to be biased and may miss important opportunities for creating successful places.

Another hypothesis from Giuliano and Small (1993) concerns equity and deserves more analysis. They suggested that racial discrimination may limit people's ability to optimize freely their job and residential locations. This is particularly relevant in regions like Detroit where racial segregation is high. Therefore this study estimated separate models for Blacks and Whites respectively. While the results show several differences in residential location choice between Blacks and Whites that suggest the existence of racial discrimination, commute time is significant for both populations. While there is no single model that can be simply applied to all geographic areas or to all demographic and social economic groups, it is important for policy makers to remember the importance of improving accessibility. The concept of accessibility and effective measurements of accessibility can provide a useful tool for other planning practice and public policies besides land use and transportation modeling. For example, environmental justice (EJ) analysis needs to identify and address any potential adverse impacts on minority and low-income population groups that could result from infrastructure investments such as transportation projects. The methodology and the results of this study could be useful for improving environmental justice analysis, particular if race-specific models are used.

The finding of the significant negative utility of commute time on residential location choice at the regional, county, and municipality levels has important implications to on-the-ground planning policies. Since commute time to work is the most important measurable factor in residential location choice, local planning policies should at least allow residential development close to employment to reduce commute time. A more proactive approach could be to encourage higher density residential development close to employment. However, many existing planning policies have the opposite effects. Master plans and zoning ordinates typically have maximum residential density limits but rarely have minimum density requirements. They also tend to emphasize separating different land uses rather than promoting land use mix. Such policies are inconsistent with people's preference on good accessibility to jobs, as well as to nonwork destinations.

Accessibility is a component of quality of life, which is a goal that land use and transportation planning try to achieve. Categorizing accessibility into regional, sub-regional, and local levels may help suggest practical solutions to improve overall accessibility. It is important to increase walkability at neighborhood level for appropriate locations. Regional planning may be most effective on large scale regional attractions with clusters of activities. At the middle range, municipalities may decide on what strategies serve them best to achieve mid-range accessibility. Do they want to intensify development and bring destinations closer to create a lively city? Do they want to "keep rural characteristics" by maintaining low density and increasing travel speed? Hopefully, competitions among communities would provide a wider range of choices to the residents so that people could "vote with their feet" in finding their preferred home locations.

D. Questions for Further Research

1. Additional Data for Spatial Cluster Analysis

In the present study, spatial clustering only shows limited success in residential location choice models. The sign of the variable is correct but its coefficient is far from being statistically significant. It only increases the model's goodness of fit slightly. While the theoretical reasoning based on trip-chaining phenomenon is sound, availability of data is more problematic. Even though the present study uses a very comprehensive and disaggregated data set, it still may not be sufficient for developing a good indicator on the degree of activity clustering. For example, a shopping center may consist of many individual retailers and service providers. However, all or most of these establishments often locate in a single building, or multiple buildings on a single land parcel. All of these establishments are represented by one single point, which is the centroid of the land parcel. In the future, a data set that can distinguish these establishments geographically may increase the accuracy of measuring the degree of spatial clustering.

2. Spatial Cluster Boundaries and Geographic Scales

Boundaries and the scale of analysis units are very important in spatial cluster analysis, for both Average Nearest Neighbor statistics and Ripley's K statistics used in the present study. A group of points may be measured as clustered if they are placed in one part of a large analysis area, but may be measured as dispersed if the area reduces to enclose those points tightly. In the present study, existing municipal boundaries or traffic analysis zones are used as boundaries to calculate cluster indices. These boundaries are artificial and vary greatly in size. A more standardized zone system is necessary in the future research to minimize boundary effects.

3. Defining Multiple Mid-ranges for Measuring Nonwork Accessibility

The mid-range accessibility concept presented in this study is based on the analysis of different characteristics between work trips and nonwork trips, as well as the concept of central place theory. The implementation of the mid-range accessibility measurement in the present study is a relatively simple cumulative opportunity measure that counts nonwork activities within a specified travel time threshold. A number of time thresholds were tested, and the 10-minute threshold was chosen because it has the most significant effects on residential location choice. However, it treats various kinds of nonwork activities equally within the time threshold. The key notion of central place theory is the hierarchy of market areas. The friction factors (Table 18) indicate that school trips are shorter than shopping trips, and shopping trips are shorter than other nonwork trips in the Detroit region, because the friction factors, i.e., the impedance to travel, decreases respectively. Travel time information from household travel survey may provide additional evidence that various nonwork activities have different scales of market areas. In future studies, multiple time thresholds may be explored for different types of nonwork activities and may be used in residential location choice models. Separating work and nonwork accessibility, as well as separating various types of nonwork accessibility based on the nature and market scales of those nonwork activities may provide a more complete explanation of accessibility's effects on residential location choice.

4. Geographic Areas for Modeling

Scale also affects sub-regional models. The present study uses a particular county and city to analyze scale effects on residential location choice. Although the county and city are carefully chosen, alternative selections of sub-regional areas are necessary to generalize the findings from this study. First, all counties and more cities should be modeled and compared to better understand the scale effects. Furthermore, it may be even better to use more standardized and uniformed sub-regional areas other than political boundaries to fully understand the effects of geographic scales on residential location choice models.

This study is based on a single metropolitan region, which is the Detroit region. More generalized conclusions on the effects of accessibility on residential location choice will depend on empirical studies of a variety of metropolitan areas, including those places that are more compact than the Detroit region and have stronger public transit systems. Public transit accessibility is not a strong component of this study, partly because the share of transit trips in the Detroit region is less than 2%.

5. Using Commute Time in Urban Modeling

This study finds that individual workers' commute time is the most influential variable in explaining residential location choice at regional and sub-regional levels. Incorporating workers' commute time into land use and transportation forecast models requires good predictions of job locations for individual workers. It requires more research and assessment. The approach in general is consistent with the general trend in the land use and transportation modeling world that is moving towards to greater disaggregation and micro simulation. Aggregated models are less productive because they suffer from ignoring individual characteristics.

E. Conclusions

This study evaluates accessibility's effects on residential location choice. In a comprehensive analysis, it considers accessibility at various geographic scales ranging from local, to mid-range, to regional levels. It incorporates both place-based accessibility measures and individual workers' commute time. It considers not only the quantities of activities but also the clustering of activities that contributes to accessibility. It does not treat accessibility as equal to all people and invariant at all scales. Instead, it analyzes accessibility for various socio-economic groups of population, particularly for population groups of various race and ethnicity, as well as at multiple levels of geography in a metropolitan region.

The disaggregated nature of the data and methodology used in this study enables the research to analyze various aspects of accessibility and their effects on residential location choice. The more refined parcel-based measures permit the detection of accessibility patterns for smaller geographic areas such as neighborhoods within walking distance, as well as incorporation of the contextual effects of accessibility at lager geographical areas including the regional and sub-regional levels.

The effects of accessibility on residential location choice are apparent in this study. The results show various aspects of accessibility matter significantly when households make residential location decisions. First and foremost, individual workers' commute time has the biggest impact on residential location choice. It explains more variation in household location choice decisions than any other variables. For two-worker households, both workers' accessibility to employment affects location choice significantly. Results of other accessibility measures are mixed in the models of the present study. There is evidence that local, mid-range, and regional accessibility measures affect residential location choice, while the effects of clustering need further research. Overall, these findings suggest that significant linkages exist between accessibility and residential location choice. It is apparent in this study that workers' personal travel time to work is the most effective accessibility measure for analyzing its impact on residential location choice. This study has demonstrated this finding to be true at multiple geographic levels in the Detroit region. The significance of this finding is that travel time to work could be more predictive in urban models while regional and local accessibility are typically used in the current practice. This study also demonstrates that the effects of accessibility on residential location choice vary for different population groups, which reflects the economic, demographic, and institutional constraints on residential location choices of various racial and ethnic population groups.

The findings of this study should help urban and regional planning policy making. Policies that emphasize amenities and services but ignore accessibility would be biased and might miss important opportunities for creating successful places. Public policies should take into account work and nonwork accessibility at various scales, although work accessibility strategies could be most effective.

BIBLIOGRAPHY

Alonso, W. (1964). *Location and Land Use*. Cambridge: Harvard University.

Banerjee, T., and Baer, W. C. (1984). *Beyond the neighborhood unit: Residential environments and public policy.* New York: Plenum.

Ben-Akiva, M., and Bowman, J.L. (1998). Integration of an activity-based model system and a residential location model. *Urban Studies,* 35 (7), 1131-1153.

Ben-Akiva, M., and Lerman, S.R. (1979). Disaggregate travel and mobility choice models and measures of accessibility. In Hensher, D.A., and Sopher, P.R. (Eds.), *Behavioural Travel Modelling* (pp. 654–679). London: Croom Helm.

Ben-Akiva, M., and Lerman, S. R. (1985). *Discrete Choice Analysis.* Cambridge, MA: MIT Press.

Burns, L.D. (1979). *Transportation, Temporal and Spatial Components of Accessibility*. Lexington: Lexington Books.

Cervero, Robert. (1996). *Paradigm Shift: From Automobility to Accessibility Planning*. (Working Paper 677, Institute of Urban and Regional Development). Berkeley: University of California.

Cervero, Robert, and Gorham, R. (1995). Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association,* 61(2), 210-225.

Cervero, Robert, and Kockelman, K. (1997). Travel demand and the three Ds: Density, diversity, and design. *Transportation Research*, *Part D,* 2(2), 199-219.

Cervero, R., Rood, T., and Appleyard, B. (1999). Tracking accessibility: employment and housing opportunities in the San Francisco Bay Area. *Environment and Planning* A, 31(7), 1259- 1278.

Christaller, Walter. (1933/1966). *Die Zentralen Orte in Suddeutschland*. Jena, Germany: Fischer, 1933. English translation by C. Baskin, *The Central Places in Southern Germany*, Englewood Cliffs, NJ: Prentice-Hall, 1966 (all text citations refer to the English translation).

Crane, R. (2000). The influence of urban form on travel: An interpretative review. *Journal of Planning Literature*, 15(1), 3-23.

Cressie, Noel. (1996). Change of support and the modifiable areal unit problem. *Geogrphical Systems, 3, 159-180.*

Dalvi, M.Q., Martin, K.M., (1976). The measurement of accessibility: some preliminary results. *Transportation*, 5, 17–42.

Domencich, Tom., and McFadden, Daniel. (1975). *Urban Travel Demand: A Behavioral Analysis.* Amsterdam: North Holland.

ESRI. (2011). *Overlapping Neighborhood Statistics: Focal Functions.* Retrieved on May 24, 2011, from http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Overlapping_neighborhood_statistics:_focal_functions

ESRI. (2012). *Average Nearest Neighbor (Spatial Statistics).* Retrieved on April 28, 2012, from http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//005p00000008000000

Frey, William, and Farley, Reynolds. (1996). Latino, Asian and Black Segregation in U.S. Metropolitan Areas: Are Multiethnic Metros Different? *Demography* 33(1): 35-50. (Available with supplementary data for individual areas as: Research Report 93-278, Ann Arbor, MI: University of Michigan, Population Studies Center).

Frey, William, and Myers, Dowell. (2005). *Racial Segregation in US Metropolitan Areas and Cities, 1990-2000: Patterns, Trends, and Explanations.* Ann Arbor, MI: University of Michigan, Population Studies Center. Retrieved on December 18, 2011, from http://www.psc.isr.umich.edu/pubs/pdf/rr05-573.pdf

Geurs, K. T., and van Wee, B. (2004). Accessibility Evaluation of Land-use and Transport Strategies: Review and Research Directions. *Journal of Transport Geography*, 12(2), 127-140.

Giuliano, G., and Small, A. K. (1993). Is the journey to work explained by urban structure? *Urban Studies*, 30, 1485–1500.

Greene, W. H. (2003). *Econometric Analysis* (5th ed.). New Jersey: Prentice-Hall.

Grengs, Joe. (2010). Job Accessibility and the Modal Mismatch in Detroit. *Journal of Transport Geography*, 18, 42-54.

Grengs, J., Levine, J., Shen, Q., and Shen, Qingyun. (2010). Intermetropolitan Comparison of Transportation Accessibility: Sorting Out Mobility and Proximity in San Francisco and Washington, DC. *Journal of Planning Education and Research*, 294, 427-443.

Hägerstrand, T. (1970). What About People In Regional Science? *People of the Regional Science Association*, 24, 7–21.

Hamilton, Bruce W. (1982). Wasteful Commuting. *Journal of Political Economy,* 90(5), 1035-1053.

Handy, Susan. (1993). Regional versus Local Accessibility: Implications for Nonwork Travel. *Transportation Research Record*, 1400, 58-66.

Handy, S. L., and Niemeier, D.A. (1997). Measuring Accessibility: An Exploration of Issues and Alternatives. *Environment and Planning A*, 29, 1175-1194.

Hansen, Walter G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners,* XXV(2), 73-76.

Hanson, S. (1980a). Spatial diversification and multipurpose travel: implications for choice theory. *Geographical Analysis*, 12, 245-57.

Hanson, S. (1980b). The Importance of the Multi-purpose Journey to Work in Urban Travel Behavior. *Transportation,* 9, 229-248.

Horner, M. W. and Murray, A. T. (2004). Spatial representation and scale impacts in transit service assessment. *Environment and Planning B,* 31(5), 785-797.

Horowitz, A. J. (1982). A comparison of socioeconomic and structural determinants of trip tour length. *Papers of the Regional Science Association* 50, 185-95.

Hsiao, S., Lu, J., Streling, J. and Weatherford, M. (1997). Use of geographic information system for analysis of transit pedestrian access. *Transportation Research Record, 1604, 50-59.*

Hu, Pat S. and Reuscher, Timothy R. (2004). Summary of Travel Trends: 2001 National Household Travel Survey. Retrieved on December 29, 2010, from http://nhts.ornl.gov/2001/pub/STT.pdf

Iacono, M., Krizek K., and El-Geneidy, A. (2008). Access to Destinations: How Close is Close Enough? Estimating Accurate Distance Decay Functions for Multiple Modes and Different Purposes. Minnesota DOT. Retrieved on July 10, 2010, from http://www.cts.umn.edu/Publications/ResearchReports/reportdetail.html?id=1610

Ingram, D. R. (1971). The Concept of Accessibility: A Search for an Operational Form. *Regional Studies,* 5, 101-107.

Kerrigan, M., and Bull, D. (1992). Measuring accessibility - a public transport accessibility index. In *Seminar B held at the PTRC Transport, Highways and Planning Summer Annual Meeting.* England: University of Manchester Institute of Science and Technology.

Krizek, Kevin J. (2003). Operationalizing Neighborhood Accessibility for Land Use–Travel Behavior Research and Regional Modeling. *Journal of Planning Education and Research,* 22, 270-287.

Krizek, Kevin J. (2005). Perspectives on Accessibility and Travel. In David Levinson and Kevin J. Krizek (Eds.), *Access to Destinations* (pp. 109-130). London, UK: Elsevier.

Kwan, Mei-Po. (1998). Space-Time and Integral Measures of Individual Accessibility: A Comparative Analysis Using a Point-based Framework. *Geographical Analysis,* 30(3), 191-216.

Leake, G.G., and Huzayyin, A.S. (1979). Accessibility Measures and Their Suitability for Use in trip Generation Models. *Traffic Engineering and Control,* 2, 566-572.

Lee, B. H. Y. (2009). Accessibility and Location Choice: Innovations in Measurement and Modeling. Ph.D. dissertation. University of Washington. Seattle, WA.

Lee, B. H. Y., Waddell, P., Wang, L., Pendyala, R. M., (2010), Reexamining the influence of work and nonwork accessibility on residential location choices with a microanalytic framework, *Environment and Planning A,* 42(4), 913 – 930.

Lenntorp, Bo. (1976). *Paths in Space-Time Environments: A Time-geographical Study of Movement Possibilities of Individuals.* Department of Geography, the Royal University of Lund.

Levine, Jonathan. (1998). Rethinking Accessibility and Jobs-Housing Balance. *Journal of the American Planning Association,* 64(2), 133-149.

Lynch, Kevin. (1962). *Site Planning.* Cambridge, MA: MIT Press.

Ma, Kang-Rae and Banister, David. (2006). Excess Commuting: A Critical Review. *Transport Reviews,* 26(6), 749-767.

Makri, Maria-Christina, and Folkesson, Carolin. (1999). Accessibility Measures for Analyses of Land-Use and Travelling with Geographical Information Systems. Presented at *Nordic Research Network on Modeling Transport, Land-Use and the Environment Conference*. Retrieved on December 27, 2010, from http://www.trafikdage.dk/td/papers/papers99/papers/paper/bpot/makri/makri.pdf.

Manning, A. (2003). The Real Thin Theory: Monopsony in Modern Labour Markets. *Labour Economics,* 10, 105–131.

Manski, Charles, and Lerman, Steven. (1977). The Estimation of Choice Probabilities from Choice-Based Samples. *Econometrica*, 45(8), 1977-1988.

Massey, Douglas, and Denton, Nancy. (1993). *American Apartheid: Segregation and the Making of the Underclass.* Cambridge, MA: Harvard University Press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.

McFadden, D. (1978). Modeling the choice of residential location, In A. Karlqvist et al. (Ed.), *Spatial Interaction Theory and Planning Models*. Amsterdam: North Holland Publishers.

McFadden D. (1981). *Econometric Models of Probabilistic Choice.* Cambridge, MA: MIT Press.

McFadden D, (1984). Econometric analysis of qualitative response models. In Z. Griliches, and M. Intriligator (Ed.), *Handbook of Econometrics, Vol. 2* (pp 1395 – 1457*)*. Amsterdam: North-Holland.

McNally, M. G., and Kulkarni, A. (1997). Assessment of influence of land use-transportation system on travel behavior. *Transportation Research Record, 1607, 105-115.*

Miller, Harvey. (2005). Place-Based versus People-Based Accessibility. In David Levinson and Kevin Krizek (Ed.), *Access to Destination*. London, UK: Elsevier.

Morris, J. M., Dumble, P. L., and Wigan, M. R. (1979). Accessibility Indicators for Transportation Planning. *Transportation Research A,* 13, 91-109.

Murray, A.T. (2001). Strategic analysis of public transport coverage. *Socio-Economic Planning Sciences,* 35(3), 175-188.

Niemeier, D. A. (1997). Accessibility: An Evaluation Using Consumer Welfare. *Transportation,* 24, 377-396.

Nishii, K., Kondo, K., and Kitamura, R. (1988). Empirical Analysis of Trip Chaining Behavior. *Transportation Research Record,* 1203, 48-59.

O'Neill, W., Ramsey, R. D., and Chou, J. (1992). Analysis of transit service areas using geographic information systems. *Transportation Research Record,* 1364, 131-138.

Peng, Z., and Dueker, K. J. (1995). Spatial data integration in route-level transit demand modeling. *Journal of the Urban and Regional Information Systems Association,* 7(1), 26-37.

Polzin, S. E., Pendyala, R. M., and Navari, S. (2002). Development of time-of-day-based transit accessibility analysis tool. *Transportation Research Record,* 1799, 35-41.

Raftery, Adrian. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology, 25, 111-163.*

Rajamani, J., Bhat, C. R., and Handy, S. L., Knaap, G., and Song, Y. (2003). Assessing 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.

Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.

SEMCOG. (2005). SEMCOG 2035 Forecast – Employment by Sector. Retrieved on August 7, 2011, from http://www.semcog.org/uploadedFiles/Programs_and_Projects/Planning/Regional_Forecast/forecasted_emp_by_sector.pdf

Shen, Q. (1998). Location characteristics of inner-city neighborhoods. *Environment and Planning B*, 25, 345–365.

Shen, Q. (2001). A Spatial Analysis of Job Openings and Access in a U.S. Metropolitan Area. *Journal of the American Planning Association*, 67(1), 53-68.

Small, A. K., and Song, S. (1992). Wasteful Commuting: A Resolution. *Journal of Political Economy*, 100, 888–898.

Southworth, Frank. (1985). Multi-destination, multi-purpose trip chaining and its implications for locational accessibility: A simulation approach. *Papers in Regional Science*, 57(1), 107-123.

Srour, I., Kockelman, K., and Dunn, T. (2002). Accessibility Indices: Connection to Residential Land Prices and Location Choices. *Transportation Research Record*, 1805, 25-34.

Steiner, R. L. (1994). Residential density and travel patterns: Review of the literature. *Transportation Research Record, 1466, 37-43.*

Taeuber, Karl, and Taeuber, Alma. (1965). *Negroes in Cities*. Chicago: Aldine Press.

U.S. Department of Labor, Bureau of Labor Statistics. (2011). BLS Handbook of Methods, Chapter 5, Employment and Wages Covered by Unemployment

Insurance. Revised on a regular basis. Retrieved on August 18, 2011 from http://stats.bls.gov/opub/hom/homch5_a.htm

U.S. Department of Transportation. (2005). Travel Model Improvement Program. Retrieved on August 7, 2011, from http://tmip.fhwa.dot.gov/content/619#synthesizer

Wachs, M., and Kumagi, T. G. (1973). Physical Accessibility as a Social Indicator. *Socio-economic Planning Science*, 7, 437-456.

Waddell, Paul, and Nourzad, Firouzeh. (2002). Incorporating Non-motorized Mode and Neighborhood Accessibility in an Integrated Land Use and Transportation Model System. *Transportation Research Record*, 1805, 119-127.

Waddell, Paul, and Ulfarsson, Gudmundur F. (2003). Accessibility and Agglomeration: Discrete-Choice Models of Employment Location by Industry Sector. Presented at the 82nd Annual Meeting of the Transportation Research Board, January 12-16, 2003, Washington, D.C. Retrieved on April 25, 2004, from http://www.urbansim.org/papers/waddell-ulfarsson-TRB03-access-agglom.pdf

Waddell, Paul, Xuan Liu, and Liming Wang. (2008). UrbanSim: An Evolving Planning Support System for Evolving Communities. In Richard Brail (Ed.), *Planning Support Systems for Cities and Regions*. Cambridge, Massachusetts: Lincoln Institute of Land Policy.

White, M. J. (1988). Urban Commuting Journeys Are Not ‘Wasteful’. *Journal of Political Economy*, 96, 1097–1110.

Wilson, A.G. (1971). A Family of Spatial Interaction Models, and Associated Developments. *Environment and Planning A*, 3(1), 1-32.

Zhao, F., Chow, L. F., Li, M. T., Ubaka, I., and Gan, A. (2003). Forecasting transit walk accessibility: Regression model alternative to buffer method. *Transportation Research Record, 1835,* 34-41.