Affordability Index - Working Paper

Preliminary Results for Discussion Purposes Only – Do not Distribute

August 10, 2005

Center for Transit-Oriented Development

Authors
Peter M. Haas, Ph.D.
with Albert Benedict, Scott Bernstein, Carrie Makarewicz, and Shelley Poticha

For more information contact:
Peter M. Haas, Ph.D., Center for Neighborhood Technology, pmh@cnt.org

CENTER FOR TRANSIT-ORIENTED DEVELOPMENT

The Center for Transit-Oriented Development (CTOD) partnership of three organizations; Reconnecting America, Center for Neighborhood Technology, and Strategic Economics. The CTOD seeks to use transit investments to spur a new wave of development that improves housing affordability and choice, revitalizes downtowns and urban and suburban neighborhoods, and provides value capture and recapture for individuals, communities and transportation agencies. The three organizations together combine expertise in market development and research, geographic analysis, urban design, standards rating and development, and strategies for urban areas.

PROJECT SUPPORT

The development of the Affordability Index is funded through a pilot effort of the Brookings Institution's Urban Markets Initiative (UMI) within their Metropolitan Policy Program. Five pilots were granted in 2004 as initial projects to help meet the goal of the UMI, which aims to "improve the quality of the information available on urban communities to unleash the full power of those markets while connecting them to the economic mainstream."
(www.Brookings.edu/metro/umi.htm).

The McKnight and Surdna Foundations provided matching funds for the Affordability Index and the Federal Transit Administration provided additional funds for a related project that contributed to the models development.

Table of Contents

Table of Figures

Table of Tables

INTRODUCTION

Background

Today, the real estate market knows how to incorporate the value of the land into the price of the home--based on its location and proximity to jobs and amenities—but there is less clarity about the accompanying transportation costs that are also a factor of a good - or bad - location. In most cases, the very same features that make the land and home more attractive, and likely more expensive per square foot, also make the transportation costs *less* expensive. Being close to jobs and commuter transit options reduces the expenses associated with daily commuting. And being in walking distance of a suburban downtown or neighborhood shopping district allows a family to replace some of the five to eight daily auto trips with one or more walking trips. Walking instead of driving reduces gasoline and auto maintenance costs, and may even allow a family to get by with one less automobile.

By contrast, in many places where single-family homes are more "affordable", or offer "more house for your money", in part because the land costs are cheaper, the transportation costs are much higher. In 2003, while the average transportation expenditures for a median income household in the U.S. was 19.1 %[1] of income, second only to housing expenditures, for half the households it was significantly higher, as much as 23%, while there are some households benefiting from costs as little as 13% to 14% of its income[2].

This information gap on the convenience factor, which we measure here as the cost of transportation associated with each place, leads to unexpected financial burdens and time constraints for households, poor location decisions by developers, missed and misplaced opportunities for municipalities, unbalanced criticism of the costs to build transit—since these critiques do not fully account for the benefits--and many other hidden costs associated with sprawl. Not only are the high costs hidden, but so are the low costs and therefore the inherent value of more convenient in-town urban and inner-suburban locations. Consequently, many of these convenient but undervalued areas often suffer from disinvestment and the ability to attract new investment and redevelopment.

---

[1] 2003 Consumer Expenditure Survey from the Bureau of Labor Statistics, http://www.bls.gov/cex/home.htm

[1] "Increasing Affordability Through Reducing the Transportation and Infrastructure Cost Burdens of Housing" and "Making the Case for Mixed Income Communities," Atlanta Neighborhood Development Partnership, Inc., May 2003, both at www.andpi.org/mici.

[2] Surface Transportation Policy Project. *"Driven to Spend: Pumping dollars out of our households and communities"*. Washington, D.C., June 2005. Available on www.transact.org.

Affordability Index Concept

The Center for Transit-Oriented Development (CTOD) seeks to reveal the hidden and unvalued information on transportation costs by developing a new measure that will model the full costs of transportation and combine it with the cost of housing in order to provide a new measure of housing affordability called the *Affordability Index (*the "Index"). This new measure will be available for each Census block group in the 35 major cities in the U.S. with fixed-rail transit and will supplement or replace the current measure of housing affordability that focuses on just the percentage of income spent on housing, e.g. 30%.

The Index will be reported as the percentage of income consumed by Housing Costs (H) plus Transportation Costs (T), as shown in the formula below. For example, for a particular Census block group, the Index may be 45% for a median household income, where 30% of income is for housing and 15% of income is for transportation.

Figure 1. Affordability Index Formula

$$\text{Affordability Index} = \frac{\text{Housing Costs (H)} + \text{Modeled Transportation Costs (T)}}{\text{Income (I)}}$$

Where Transportation Costs include the modeled cost of Auto Ownership, Auto Use, and Transit Use

We expect this new measure will be used by the following market actors:

- The U. S. Department of Housing and Urban Development (HUD) to supplement or replace its current measures of housing affordability
- Secondary market actors and housing lenders to give credit to location decisions that reduce transportation costs and free up income for housing,
- Housing providers and planners as a location decision tool, and

[3] 2003 Consumer Expenditure Survey from the Bureau of Labor Statistics, http://www.bls.gov/cex/home.htm
[3] "Increasing Affordability Through Reducing the Transportation and Infrastructure Cost Burdens of Housing" and "Making the Case for Mixed Income Communities," Atlanta Neighborhood Development Partnership, Inc., May 2003, both at www.andpi.org/mici.
[4] Surface Transportation Policy Project. *"Driven to Spend: Pumping dollars out of our households and communities"*. Washington, D.C., June 2005. Available on www.transact.org.

- Households to help find housing and develop household budgets.

Model Overview

To reliably model an accurate and fine-grained index, the model must account for the various characteristics specific to a location that influence transportation costs. It must also control for certain household characteristics that also determine transportation costs, somewhat independently of location, such as household income and household size. Therefore, the Affordability Index formula incorporates a set of independent variables that represent the relevant *local environment* and *household* characteristics that each influence the dependent variable - household expenditures on transportation.

Additionally, the independent variables, which co-vary and are completely interdependent on each other, are combined to calculate transportation costs in three separate components: auto ownership, auto use, and transit costs. These three dependent transportation costs are then summed to attain total household transportation costs. Total transportation costs are then added to actual reported housing sales costs or rents. It should be noted that the Index only models transportation costs. It is not necessary to model housing costs because, unlike transportation costs, actual housing costs are more readily available, through posted sales and rents.

The literature on transportation costs is quite large. We drew from this literature to select relevant independent local environment and household characteristics that influence auto ownership, auto use, and transit use[5]. The table below provides a detailed description of each of the independent variables or measures used in the model.

[5] John Holtzclaw, Robert Clear, Hank Dittmar, David Goldstein and Peter Haas. "Location Efficiency: Neighborhood and Socio-Economic Characteristics Determine Auto Ownership and Use---Studies in Chicago, Los Angeles and San Francisco", Transportation Planning and Technology, Vol. 25(1),pp 1-27, March 2002.

Table 1. Independent Variables: Local Environment and Household Characteristics		
Variable	**Source**	**Model Use**
Household Characteristics		
Household Income	Census 2000	Influences auto ownership and use
Household Size	Census 2000	Influences auto ownership and use
Local Environment Characteristics		
Households per residential acre	Census 2000	Provides a measure of density which influences auto ownership and use
Households per total acre	Census 2000	Provides a measure of density which influences auto ownership and use
Average block size in acres	Census/ TIGER/Line®	Block size contributes to walkability of the area, which influences auto ownership and transit use
Transit Connectivity Index	FTA 1995 Bus Routes Transit Database, local transit agency system data	Availability and extent of transit influences transit use
Distance to Employment Centers	Census Transportation Planning Package (CTPP) 2000	Distance to nearby jobs influences auto ownership and auto use
Job Density- Number of jobs per square mile	All jobs and locations from CTPP 2000	Number of nearby jobs influences probability of working at the nearby employment center
Access to amenities	Service jobs in the CTPP 2000	Access to nearby services in walking distance influences auto use and ownership, as well as transit availability and use

These independent variables were then used to model the three dependent variables in the Transportation Costs model.

Table 2. Dependent Variables: Transportation Costs	
Variable	**Source**
Auto Ownership (vehicles per household)	Modeled from independent household and local environment variables
Auto Use (annual miles driven per household)	Modeled using the 2001 NHTS reported VMT fitted to the independent variables
Transit Rides per day	Modeled from independent household and local environment variables

Limitations

We feel the Affordability Index, to the extent possible, models the average expected transportation costs affiliated with a given place, for the given household parameters. Additionally, since it is a model, the inputs can be updated and replaced in order to improve the outputs. In this run, we use data from 1999 to 2001, since the best block group data available is from the 2000 Census. While newer years are available for other sources, we chose to use the 1999 HMDA and the 20001 NHTS, to correspond to the Census.

However, although new inputs will improve the outputs over time, this is a social science model with the objective to create a standardized and national tool that reliable and cost-effectively predicts costs that are determined in part by household behavior. Therefore, we do not attest to predict every unique household's actual transportation costs for every single geographic area. For example, the model will not as reliably predict the transportation costs for households with out of the ordinary situations, such as a household having several children of driving age with each driving to separate schools and jobs, or of a very wealthy three-person household owning zero cars. This is because the model works primarily on averages and cannot account for unique or exceptional household decisions or outside the standard deviation. For these cases, not in the middle of the distribution, there are simply fewer cases in these instances from which to accurately develop a model that would also predict their transportation costs.

The other two limitations in the model are the date of the available data inputs, and the quality or availability of data. We address each of these below.

Data Timeliness. We have developed the model for a snapshot of affordability in 2000 since the majority of data at the level of geography needed for the model is from the 2000 U.S. Census. Therefore, as we move further from the 2000 U.S. Census, the accuracy of the Affordability Index will diminish for some neighborhoods, particularly those that are rapidly changing, either in terms of the local environment or the types of households that live there. Already, the 2000 Census does not reflect the new Hiawatha Light Rail Line in the Minneapolis-St. Paul Area, nor the resulting development that has occurred along the line. Until we are able to use the American Community Survey (ACS) to updated the demographic data at the tract or block group level, the data in the Index will be less accurate as we move further from the

Decennial Census. However, when the ACS is available, we could use it to update some of the demographic variables in the model.[6]

Certain variables in the Index can be updated more regularly: the Transit Connectivity Index can be updated using the local transit agency files to the extent they are updated, household density may be updated where local MPOs have accurate measures of development; annual housing prices are available from the most recent year of HMDA data, we currently use 1999 to correspond to the 2000 Census; and the distance to jobs and access to amenities may possibly be updated annually by using the new Local Employment Dynamics (LED) data from the Longitudinal Employer-Household Dynamics program of the Census. This new data set will combines the state level ES202 data with the Census Long Form, and its eventual replacement, the American Community Survey (ACS). The LED data are not currently used in the Affordability Index, since the program is in pilot mode and is not yet publicly available for the initial pilot state, Minnesota, or the other 12 pilot states. When the data are released, however, we plan to use the data on employment type, which is available at the 2-digit NAICS code, to improve our measure of "access to amenities". We currently represent access to amenities with service sector employment from the CTPP 2000 data which only goes to the 1-digit NAICS code. We expect the Minnesota state data to be ready by September 2005 with the other states following throughout the year.

Data Availability and Quality. Three variables in the model could have more extensive actual data on which to base the model; the location of transit bus stops, the number of miles driven annually per household by location, and the number of transit rides per household by location. In the current model, we do not incorporate bus stop locations as they are not in the Metro GIS file. Bus stops would give us a better measure of transit service levels for the block group since a bus line without a stop does not equate to transit accessibility. We derived the annual vehicle miles traveled from the actual miles reported by respondents in the 2001 National Household Travel Survey which we then paired to the actual block groups in which those households lived for the block groups we could unquestionably identify based on the information in the survey data. However, this diminished our sample size from approximately 60,000 records with household VMT in the NHTS to 6,840 records. Another possible source for VMT, which is similar to our interpretation of the NHTS data, is Oak Ridge National Laboratory's work on a the

---

[6] The ACS is an ongoing survey that the census bureau plans to use in place of the decennial census long form at a five year interval. Although the details of the ACS are still under review, the census hopes to address the needs of its users by providing estimate based multi-year measurements in place of those based on the current static long form.

transferability model using the Nationwide Personal Transportation Survey (NPTS). They are estimating driving characteristics at a census tract level using a cluster analysis. CTOD will explore the work that Oak Ridge is undertaking with the NPTS and compare it to our estimate of VMT with the NHTS once their model is available. Ultimately, rather than a sample and survey, actual State vehicle emissions testing programs, which measure miles driven by all registered vehicles in a state and track that back to an address, could provide a much more accurate measure of vehicle miles driven. Minnesota has canceled their emissions testing program, and other states are not willing to remit the data. It is also a huge data set that is difficult and time consuming to use.

Finally, the number of transit rides per household would be more accurate if trip origins and destinations were reported by the transit agency at the block group level. Instead, we assumed transit use was based on use of transit for the work commute as reported in the journey to work section of the Census. This misses households who use transit for other trips but not for the work commute, or vise versa, or that do not have a work commute.

Despite the limitations, we believe the Affordability Index model is a solid model framework for providing more information on the average transportation costs associated with a small geography While it is currently based on 2000 data, going forward, the model output will continue to improve as it is updated with not only more recent information from the same data sources but also with new information sources that are not currently available, such as LED, or better transit data.

METHODOLOGY

Model Description

In this section we will explain both the dependent variables and the independent variables. The transportation costs are limited to the cost of auto ownership, auto usages, and transit costs. We do not attempt in this analysis to estimate the cost of time. These three transportation costs, or components, are first modeled separately and then summed to get the total average household transportation costs.

Our analysis shows that location is important in modeling the cost of transportation. To capture this in the Affordability Index, we broke the independent variables in to two classes:

- local environment variables; and
- household variables.

We define "local environment" variables as those that are geographic in nature and reflect the built and economic environment; they include urban form, transportation connectivity, access to jobs, and walking access to amenities. The household variables are those that are intrinsic to the household itself and include in this analysis household income and size. In order to be able to separate out the geographical effects from the household or consumer effects we have structured the equations to estimate each of the transportation cost components in the following way:

$$T_{cost} = C_{tsp} * F(X_{le}) * G(X_{hh})$$

Where T_{cost} is the total cost of that particular transportation cost (ownership, use or transit ride); C_{tsp} is the cost per unit of that entity (e.g. dollars per transit ride); X_{le} are the local environment variables; and X_{hh} are the household variables. We fit for $F(X_{le})$ first so that it in itself is an estimate of the entity, and then refine that by fitting for $G(X_{hh})$ while holding $F(X_{le})$ constant, therefore $G(X_{hh})$ can be thought of as a tweak to the entity estimate correcting for the household. We derived the per unit costs (C_{tsp}) for each transportation component from available data, where available, and from the literature on standard costs for these components.

Model Inputs

The affordability index uses nationally available public datasets – most of which are available for download over the Internet; U.S. Decennial Census, U.S. Census TIGER/Line® Files, Census Transportation Planning Package 2000, National Transportation Atlas Database, FTA Bus Routes Database, 2001 National Household Travel Survey, 1999 Home Mortgage Disclosure Act (HMDA) data, and information on the metropolitan area's transit system lines and service where available. The Minneapolis-St. Paul region is above average in terms of making available local GIS datasets – the local transit data, and other Census products for the region were obtained from MetroGIS. Once acquired, the data were incorporated into a GIS application for aggregation and spatial analysis and then into SPSS for statistical analysis. The data were gathered at either a census block group, where available, or at census tract level if block group data was not available. Following is a description of the variables, their source, and how they combine to create the Affordability Index.

1. Household Demographics (Census 2000)
2. Walkability and Access to Amenities (U.S. Tiger Files & CTPP 2000). Note: In future releases, this data may be replaced with the Longitudinal Employment Household Data from the Census as it becomes available for each state.
3. Location of jobs (CTPP 2000)
4. Transit Data (1995 FTA Bus Routes Database/Current Transit Agency system data)
5. Auto Usage (2001 NHTS)
6. Auto Costs (2001 Federal Highway Administration Car Cost Guide)
7. Housing Costs (1999 HMDA loan data & Census 2000)

Household Demographics and Socioeconomic Characteristics

Source:
U.S. Decennial Census: www.census.gov.

Variables:

- Block datasets are from the Census 2000 Summary File 1 (SF 1) 100-Percent Data and include Census variables on population (P1) and total households (P15).
- Block Group datasets are from the Census 2000 Summary File 3 (SF 3) - Sample Data on Aggregate Number of Vehicles Available by Tenure (H46) to calculate auto ownership, Median Household Income in1999 Dollars (P53) and Means of Transportation to Work for Workers 16 Years and Over (P30).

Use

The census data were used for both household and local environment variables, including GIS files.

Walkability and Access to Amenities

Source:

- GIS TIGER/Line® : www.geographynetwork.com or www.census.gov/geo/www/cob/.
- Census Transportation Planning Package (CTPP 2000): http://www.fhwa.dot.gov/ctpp

Variables:

- GIS census block file (Tiger Line File)
- Total number of jobs & number of service jobs (CTPP 2000)

Use

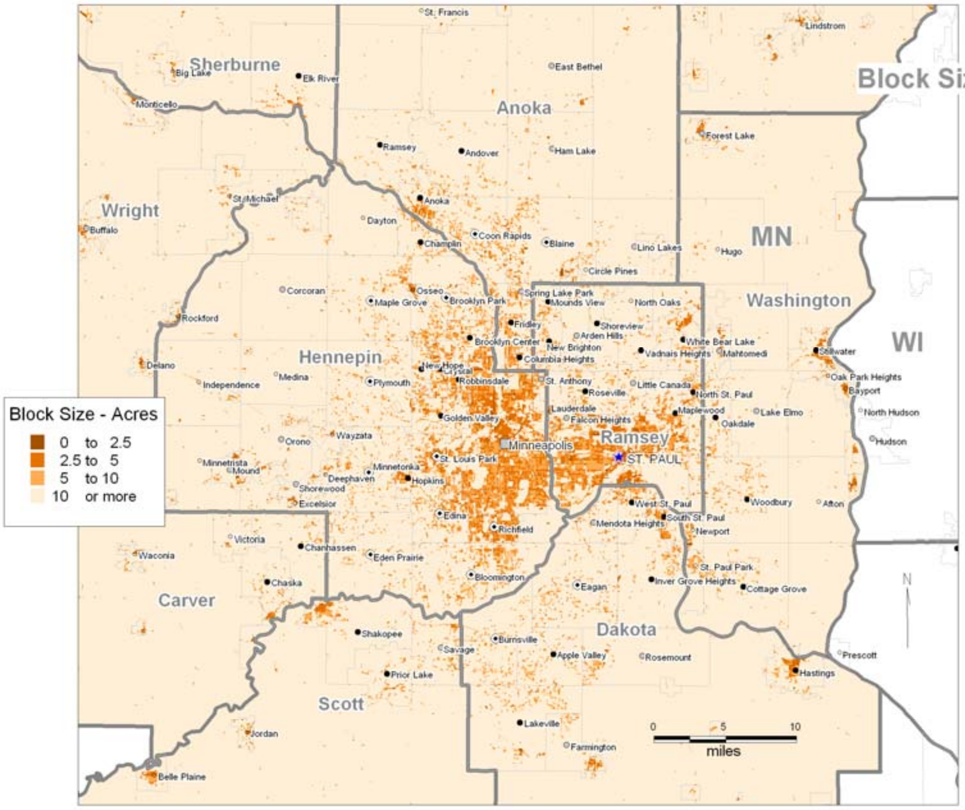
Two proxies are used to assess the walkability of an area: 1) the average Block size, and 2) the total number of service sector jobs.

1. Average Block Size: We calculate the Average Block Size using the Census TIGER/Line® files by dividing the total acres of a Block Group (BG_acres) by the number of the Census Blocks within that particular Block Group (B_count) as follows:

   $$\text{Average Block Size} = \text{BG\_acres} / \text{B\_count}$$

   A smaller average block size indicates interconnected streets, which generally implies an interconnected neighborhood with shorter blocks and narrower streets. These characteristics, in turn, lead to slower traffic and fewer lanes of traffic to cross making the blocks more “walkable”. Figure 2 below depicts the range of Average Block Size for the Minneapolis/St. Paul region. As the map indicates, the blocks in the cities of Minneapolis and St. Paul are smaller and typically more walkable than the rest of the region

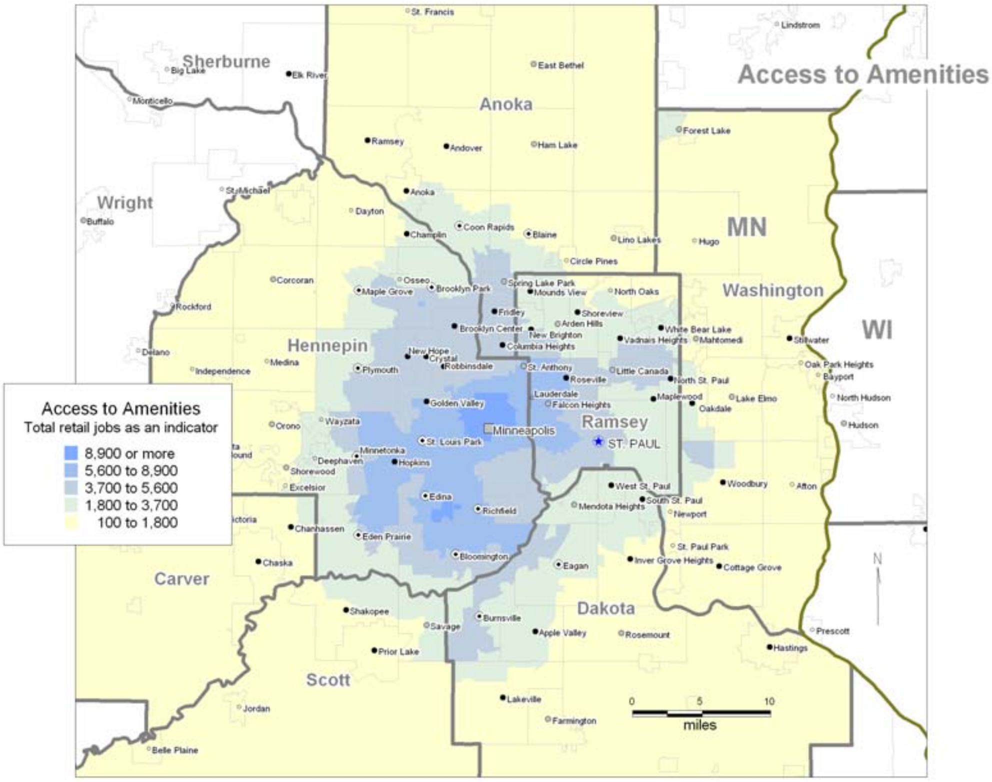
Figure 2. Average Block Size



Source: U.S. Census Bureau, TIGER/Line Files

2. Access to Amenities is a proxy of the number of nearby service jobs are derived. Using Table 2 from the 2000 Census Transportation Planning Package we calculated the number of service sector jobs/square mile and proportionally estimated them to the Census Tract. Figure 3 depicts the number of service sector jobs in the Minneapolis/St. Paul region by Census Tract. The map below parallels the block size map in that areas with a denser street grid also have a greater density of nearby service and retail amenities.

Figure 3. Access to Amenities



Source: The Census Transportation Planning Package (CTPP) 2000

Employment Centers

Source:

U.S. Census Transportation Planning Package 2000 Table 2 (CTPP 2000) www.fhwa.dot.gov/ctpp

Variables:

The CTPP 2000 dataset is a partnership between the State Departments of Transportation and the American Association of State Highway and Transportation Officials (AASHTO). The CTPP 2000 is a set of special Census tabulations to specifically support transportation planning. The CTPP 2000 summarizes Census Person information by place of residence, place of work, and worker-flows between home and work. For the Affordability Index, it is used to identify the location, size, and job density of employment centers in each region.

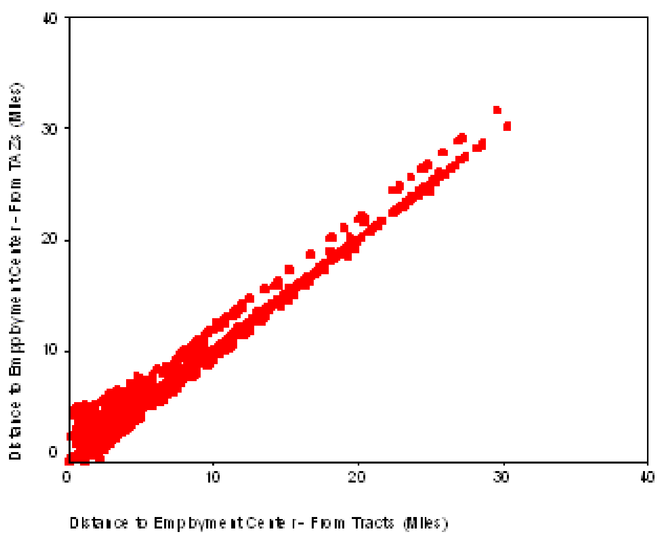
Use

In the Affordability Index model, Employment Centers serve as a proxy for the distance a household may travel to work, and therefore, how many annual vehicle miles they may travel as well as how many vehicles the household may own. These two factors are two major portions of total household annual transportation expenditures.

However, it is important to note that the journey to work is not the only predictor of auto ownership and transportation costs. While higher mileage increases total transportation expenditures, the number of autos has a more significant impact on expenditures. And the number of autos owned may be more a factor of the walkability and access to amenities of a place than is the distance to work. According to the 2003 National Household Travel Survey, the average household makes 5-8 trips per day of which only one is the trip to work. Therefore, while distance to Employment Centers is an important factor within the Index, it is complimented by the walkability, access to amenities, and the Transit Connectivity Index.

To locate and define the size of the employment centers for a region, we use the CTPP 2000 package that provides the total number of employees per census tract. Although the CTPP 2000 provides the employment data at three scales appropriate for this analysis; block groups, TAZ's and census tracts, we used the census tract files since they had the most complete geographic coverage within the CTPP. Although the CTPP publicizes Block Group geography data, this is only true for a select number of counties in each state, and in Minnesota, for example, the pilot area counties are not included. To ensure we were not loosing information or significantly impacting the outcome by using the Census Tracts instead of the Transportation Analysis Zones, we compared the two through a regression analysis. As Figure 6 shows, they are highly correlated.

Figure 4. Correlation between employment centers with TAZ and Census Tracts

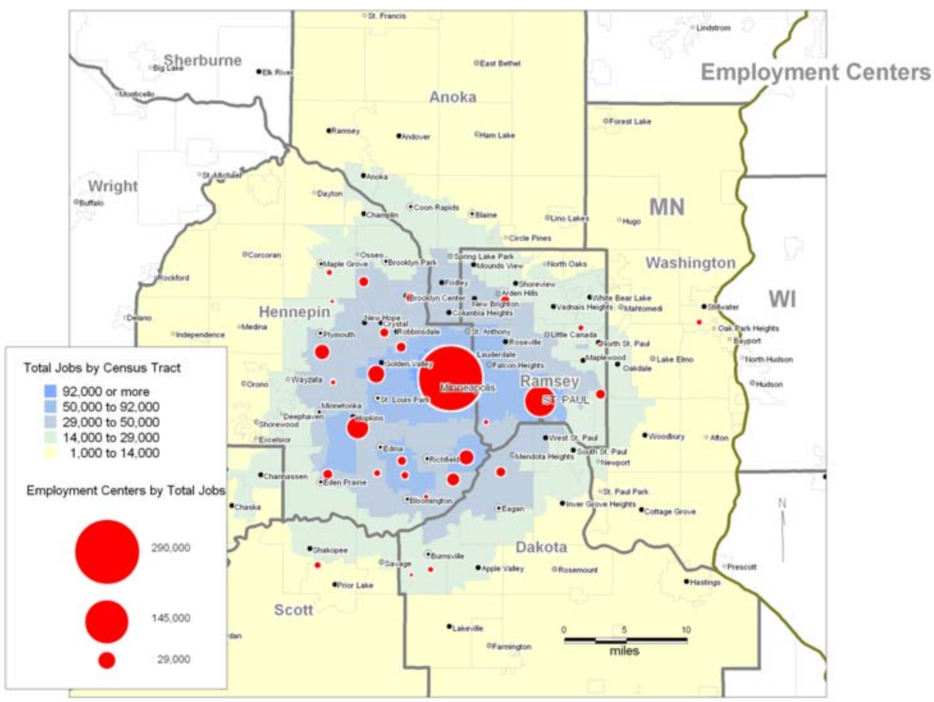


Employment Center Analysis

This analysis used a simple clustering analysis to determine where the centers of employment are within the region and the size of each employment center based on the number of employees within its boundaries.

The following map shows the results of the employment center cluster analysis. The shading indicates the total number of jobs/square mile and the symbol proportionally represents the location of the geographic location of the employment center by the total number of employees.

Figure 5. Location and Size of Employment Centers



Source: The Census Transportation Planning Package (CTPP) 2000

To determine the significance of the employment centers for a given block group, we use two measures. First, the simple distance to the nearest large employment center and second the sum of the ratio of the size of the employment center over the distance squares - this gravity model measures the strength or pull of all the centers to the specific location. We assess the relative importance of these two measures in our regression analysis.

Transit Data

Source:

- National Transportation Atlas Database (NTAD 2003) of fixed rail stations + CTOD National Transit Database.

- The 1995 FTA Bus Route GIS Database is a national bus route database with GIS files created through a collaborative effort between the FTA and Bridgewater State College (BSC) in 1995. The database provides information for 550 communities in the United States. Available from *http://geolab.bridgew.edu/docs/busroutes*

- Current GIS bus files and level of service data were gathered from MetroGIS datafinder website (http://www.metrogis.org/). MetroGIS is a collaborative effort among more than 250 municipalities in the Minneapolis/St. Paul area to share and make publicly available geospatial data (http://www.metrogis.org/).

Variables:

- GIS bus route database with Level of Service (LOS) information

Use

The bus data were used to calculate the availability, or connectivity of transit in a given location, referred to as the Transit Connectivity Index (TCI)[7]. The TCI was calculated using 1995 and the current Minneapolis bus route data so the differences could be compared and their impact on the model quantified. As the pilot region for the Affordability Index it was necessary for us to calibrate the model using both the current and 1995 datasets, since many metro areas will not have a current local GIS bus system to utilize and therefore we need to know the impact of using the older national 1995 database.

The TCI measures the quality of bus and fixed guideway transit service, in terms of frequency and extent of service. We do this by measuring the service area of bus and fixed rail lines, their proximity and connectivity to one another, and their frequency at the Census Block Group level. We are then able to assign a transit connectivity score to each Block Group. The more bus routes running through a block group and the more often they run, the higher the score will be for that block group.

Data for the TCI

To calibrate the TCI portion of the Index we are performing two runs of the TCI in the Minneapolis/St. Paul region; each with a different data source. In one run we use the current Minneapolis bus system data available from MetroGIS. In the other run, we are using the bus system data from the 1995 FTA Bus Route Data and Bridgewater State College. For fixed guideway transit, we will always use the NTAD & the CTOD database, however, in Minneapolis/St. Paul, the fixed guideway was not open until fall 2004 and therefore we are not

[7] The Hiawatha Line was not used in the TCI calculation since the index is based on data collected prior to the development of the system.

using train line or station data for this pilot area since the Census Household travel data is for 1999.

The purpose of the two runs is to test the quality of the 1995 Bus System Data. Preferably, we will be able to use the 1995 Bus System Data in the national Affordability Index model since it is an already assembled national database. However, if the resulting TCI scores from the two runs in Minneapolis are substantially different, we will gather the bus system files from each region where they are available.

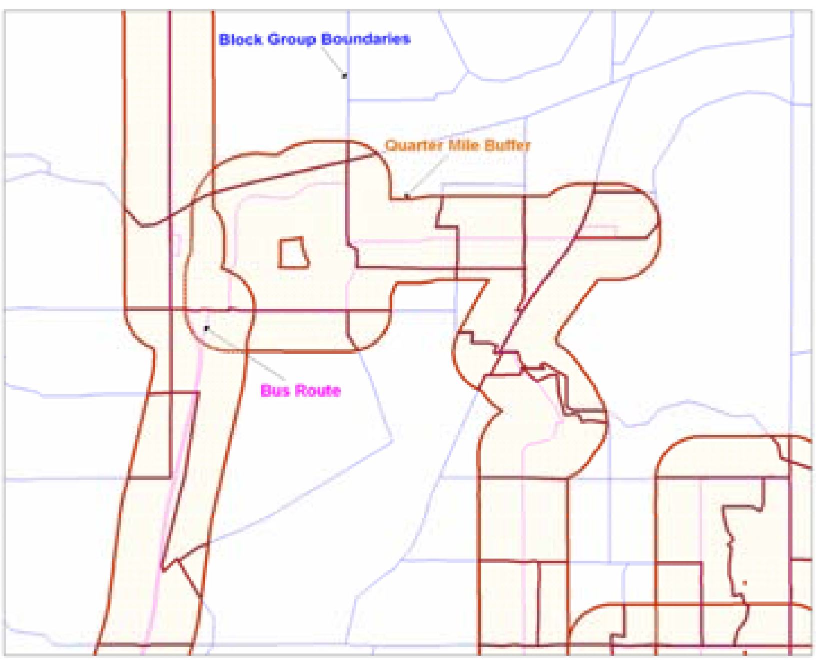
The Block Group is the geographic base of analysis in the TCI.

TCI Analysis

The accessibility portion of the TCI assumes people can access the transit within a quarter mile around both sides of the bus route and within a half mile of a fixed guideway transit station. These buffer areas capture all the block groups that may be served by a bus route, even if the route is not directly in that block group.

Figure 6 below is an example of the quarter mile buffer around the bus routes in relation to block group boundaries.

Figure 6. Quarter mile buffer around bus routes in relation to block group boundaries



Given the service area, the *quality* of service is then measured by combining the total weekly trips for each bus or train route--regardless of direction, location or pattern, in the buffer area.

To combine the service area with the quality of service the quarter mile buffer is split by the Block Groups and the split area is multiplied by the total trips of the route. The area by total trips is then summarized to the Block Group and divided by the total area of the Block Group. This calculation yields the number of rides the average person within the Block Groups has access to within a quarter mile of a bus route(s).

$$\text{TCI} = (\Sigma((\tfrac{1}{4} \text{ mile buffer area}) * (\text{route total trips}))) / \text{total block group area}$$

Figures 7 & 8 shows the resulting TCI at the block group level for the Minneapolis/St. Paul region – based on current and 1995 bus route LOS.

Figure 7. Transit Connectivity Index calculated with the current Transit System Database from the local Transit Agency

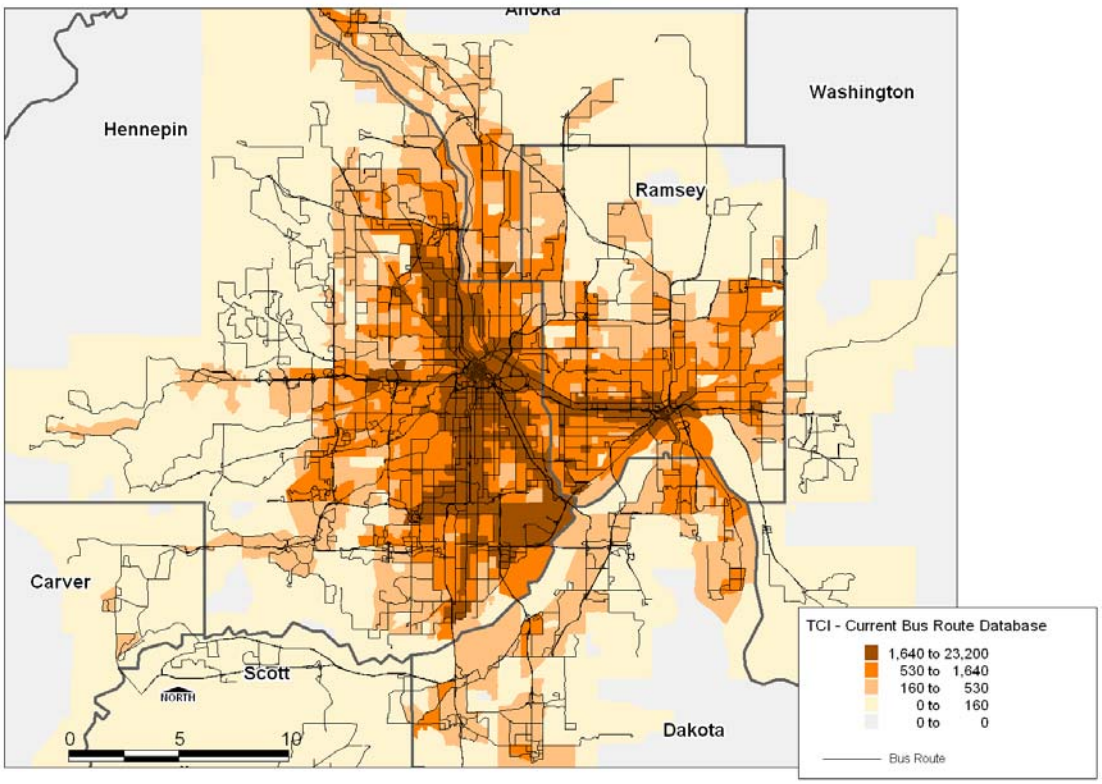
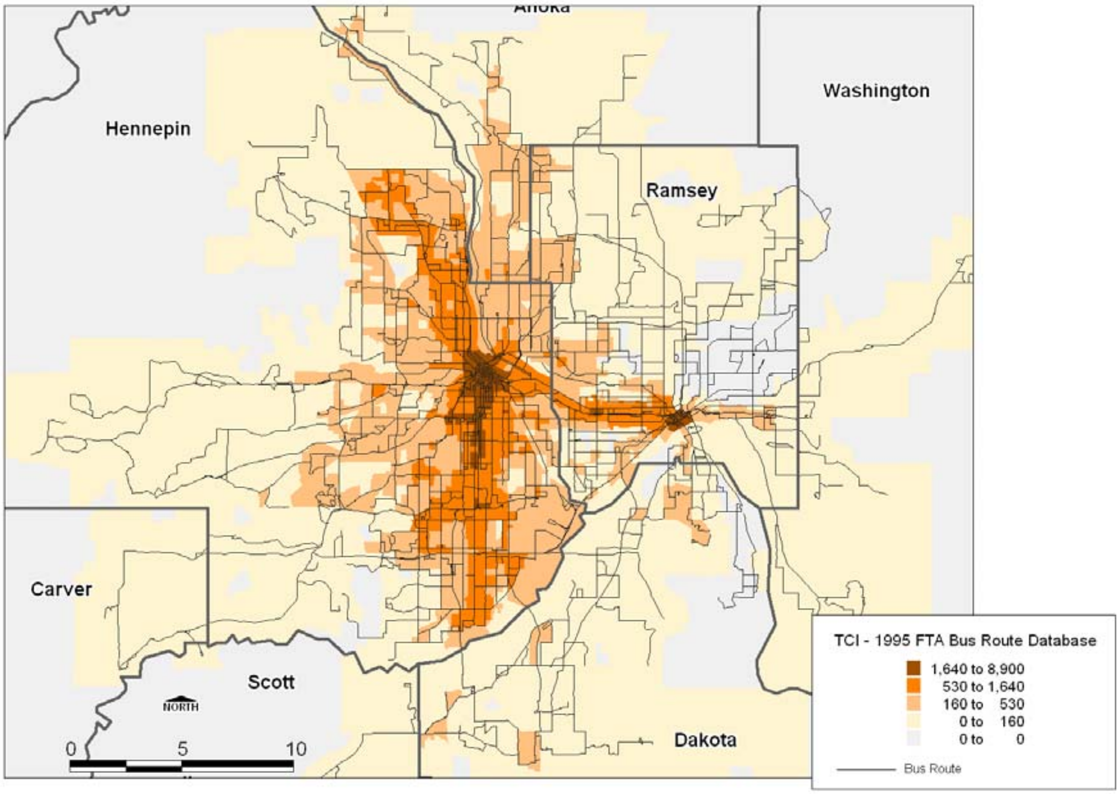


Figure 8. Transit Connectivity Index calculated with the 1995 FTA Bus Route Database



The TCI maps of the region shows what one would expect: the central cities of Minneapolis and St. Paul have the highest TCI; the TCI is highest along bus routes and especially where there is a concentration of multiple bus routes; and the TCI decreases as one travels outward to the suburban fringe areas. The above maps pertain to the transit data gathered using the metroGIS website and 1995 FTA Bus Database. A more formal statistical comparison needs to be performed to determine the impact of each data source on the model.

Auto Usage

Source:

- National Household Travel Survey (NHTS) - Bureau of Transportation Statistics – www.bts.gov
- Federal Highway Administration (FHWA) Cost of Auto Transportation Calculator/Study

Variables:

- Extrapolated at a block group level total number of vehicles and VMT per household (NHTS)

- Table illustrating the costs of owning a vehicle (FHWA)

Use:

We used VMT to estimate auto use and then the FHWA formula to determine the cost of owning and operating a vehicle.

After investigating several sources on actual driving patterns at a census block group or tract level we realized we could use the NHTS to model annual vehicle miles at the household level for a specific small scale geography. The NHTS collects data on vehicle ownership and VMT by auto and reports those data at a county level, but also includes six other characteristics of the block group that are somewhat unique to the reported block group. By matching these seven reported fields to those same fields in the Census, such as percent rentals, household and population density, as well as County and State, we were able to match 6,840 survey records with the actual census block group. This gave us a significant sample of VMT by households by block group. For a more in-depth discussion on how we used the NHTS to develop a large sample of household VMT, see the Modeling Section on Auto Use beginning on page 53.

Auto Costs

Source

Federal Highway Administration estimates based on the 2001 editions of "The Complete Car Cost Guide" and "Complete Small Truck Guide" from Intellichoice, Inc. and sales figures from "Automotive News."

Transportation Energy Data Book 2005 provides an estimate of the U.S. fleet of passenger vehicles by type currently in operation.

Use

We used the average figures in this guide for the per unit cost of auto ownership and auto use. The table below summarizes the fixed and variables costs of various passenger automobiles and then creates a weighted average based on the percentage of each vehicle type currently on the road today.

Vehicle Stock and New Sales in the US 2002 calendar year Transportation Energy Data Book - Table 3.4						
	Type	*Number*	*Percent of Total Fleet*	*FHWA Category*	*FHWA Annual Ownership Cost*	*FHWA Usage Cost Mile*
Autos	two seater	1,889	1%	Subcompact	$3,606	$ 0.06
	minicompact	1,153	1%	Subcompact	$3,606	$ 0.06
	subcompact	24,556	11%	Subcompact	$3,606	$ 0.06
	compact	43,337	20%	Compact	$4,738	$ 0.08
	midsize	40,685	18%	Intermediate	$5,250	$ 0.09
59%	large	19,452	9%	Full Size	$5,723	$ 0.10
131,072	small pick up	13,348	6%	Compact Pick Up	$4,502	$ 0.08
Trucks	large pick up	22,174	10%	Full Size Pick Up	$5,342	$ 0.10
less than 8500 lbs	small van	13,758	6%	Mini Van	$5,678	$ 0.10
	large van	5,407	2%	Full Size Van	$5,824	$ 0.10
	small suv	6,592	3%	Compact Utility	$5,107	$ 0.09
	medium suv	13,313	6%	Intermediate Utility	$5,757	$ 0.10
	large suv	3,958	2%	Full Size Utility	$5,925	$ 0.11
trucks	pick up	4,898	2%	Full Size Utility	$5,925	$ 0.11
8500 -- 10000 lbs	van / suv	1,748	1%	Full Size Utility	$5,925	$ 0.11
41%	10000-26000 lbs	2,734	1%	Full Size Utility	$5,925	$ 0.11
91,119	2600 lbs & over	3,189	1%	Full Size Utility	$5,925	$ 0.11
222,191	Autos & Trucks		100%		Weighted Average	Weighted Average
					$5,068.46	$0.090

Housing Costs

Source:

- Federal Financial Institutions Examination Council (FFIEC) - www.ffiec.gov
- U.S. Census Bureau – www.census.gov
- Federal Housing Finance Board - www.fhfb.gov

Variables:

- Home Mortgage Disclosure Act data (HMDA) – FFIEC: HMDA data on originated home purchase conventional loans at a census tract level
- Owner Occupied Aggregate Housing Value H86 & Total Owner Occupied Housing Units from 2000 Census
- Information on the average term, rate, and down payment by year and metropolitan area - Federal Housing Finance Board

Use

To determine the housing of the H + T equation we used a combination of HMDA and Census Bureau data on housing costs. The advantage of the HMDA data is that it provides a record of the actual loan amount at a census tract level. This information, coupled with the average term, rate and down payment information from the Finance Board's website, we are able to calculate the actual costs of purchasing a home in a given area. For census tracts where there were no home purchase loans listed in the HMDA database we used the Owner Occupied Aggregate Housing Value (SF3, H86) divided by the total number of Owner Occupied Housing Units to calculate the self-reported average housing value at a census tract level.

The maps below show the difference in housing values between the 1999 HMDA data on originated conventional home purchase loans versus census data on the average housing value. The map shows that people in Minneapolis and St. Paul typically underestimate the value of their property when compared to the average selling price of homes in the same census tract.

Figure 9. Average Housing Values as reported in the 2000 U.S. Census

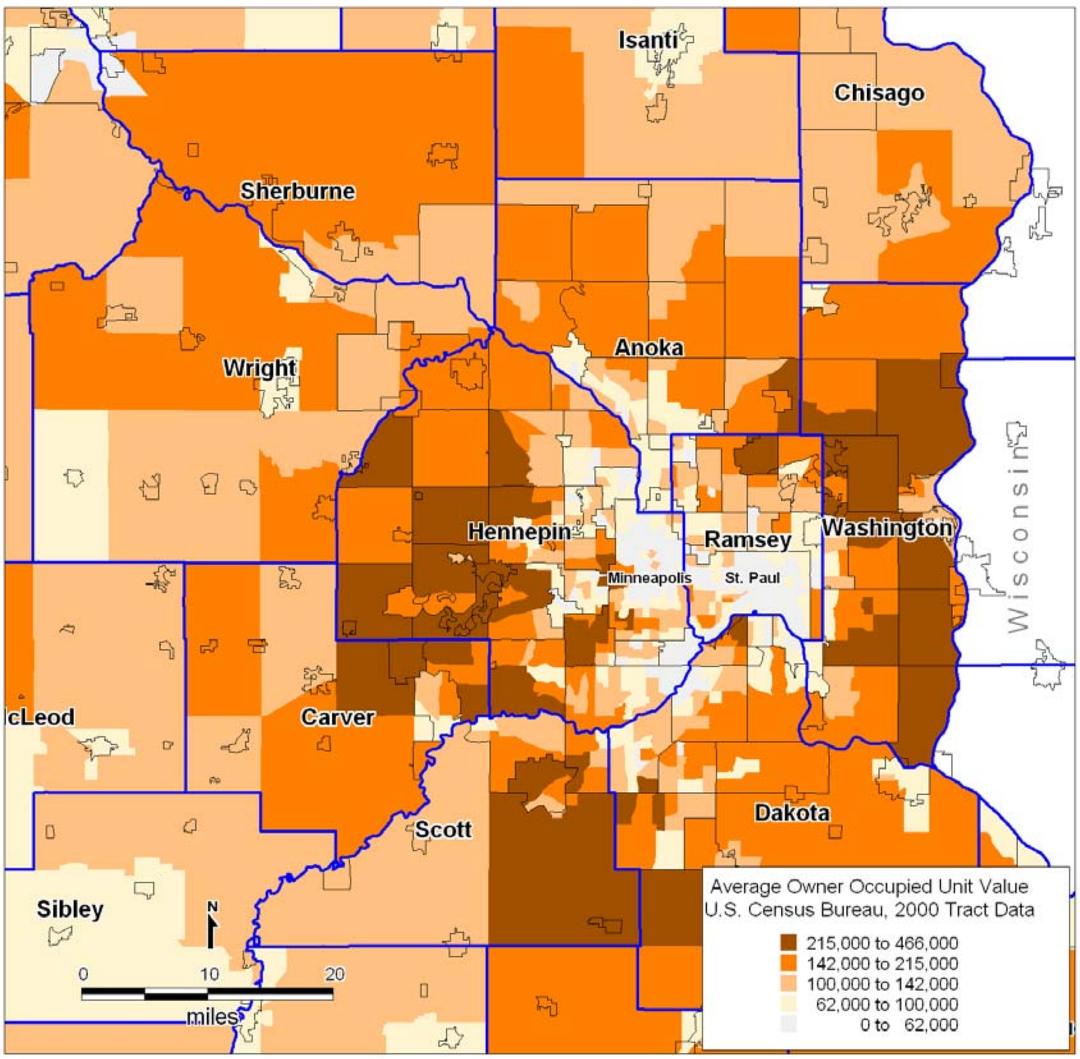
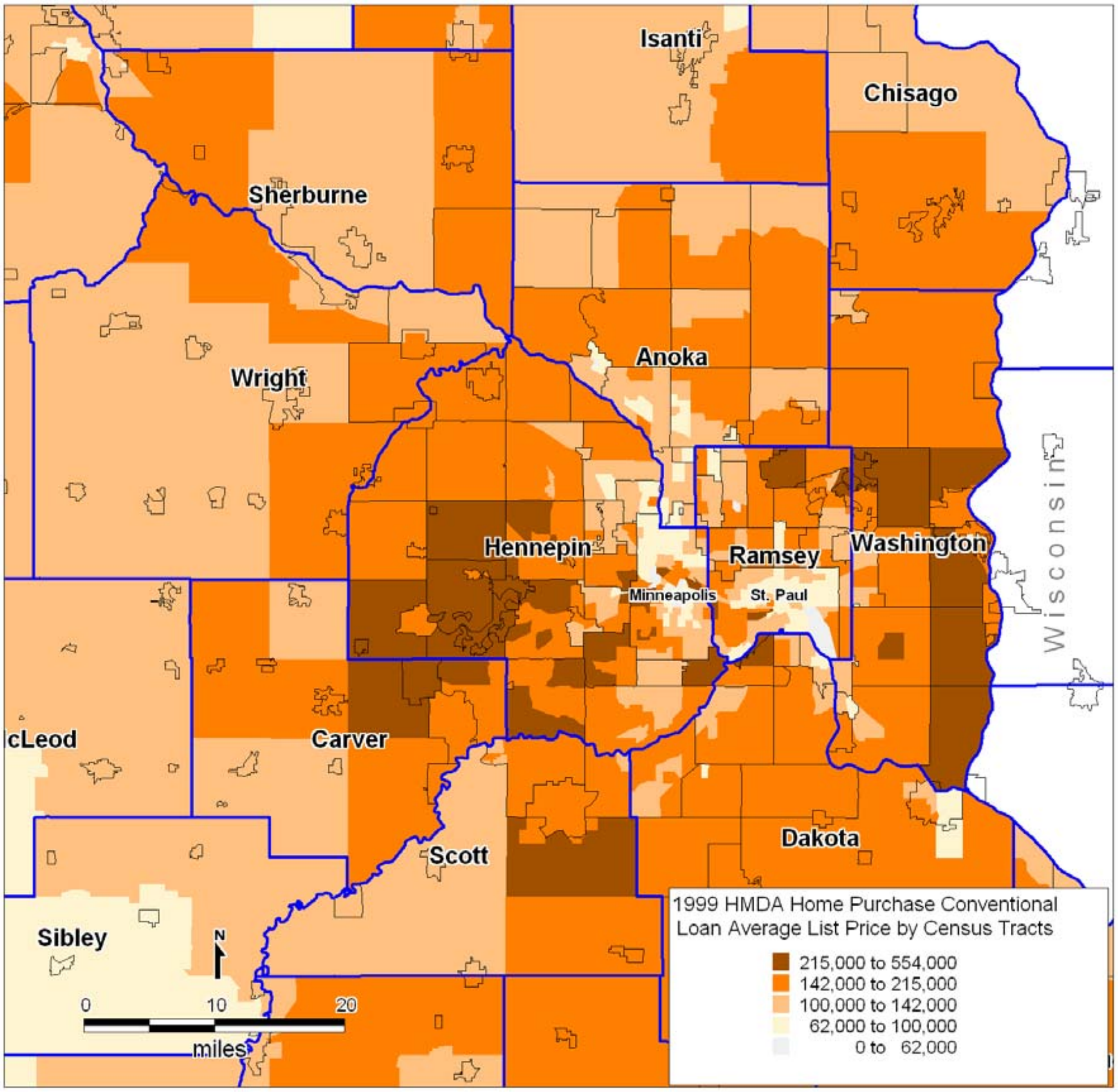


Figure 10. Average Housing Values calculated from 1999 HMDA Home Purchase Records



Model Development

How location and household attributes combine to affect a household's transportation costs is the essence of this modeling effort. This section describes the formulation of the model for each of the three components of transportation cost; auto ownership, auto use, and transit use.

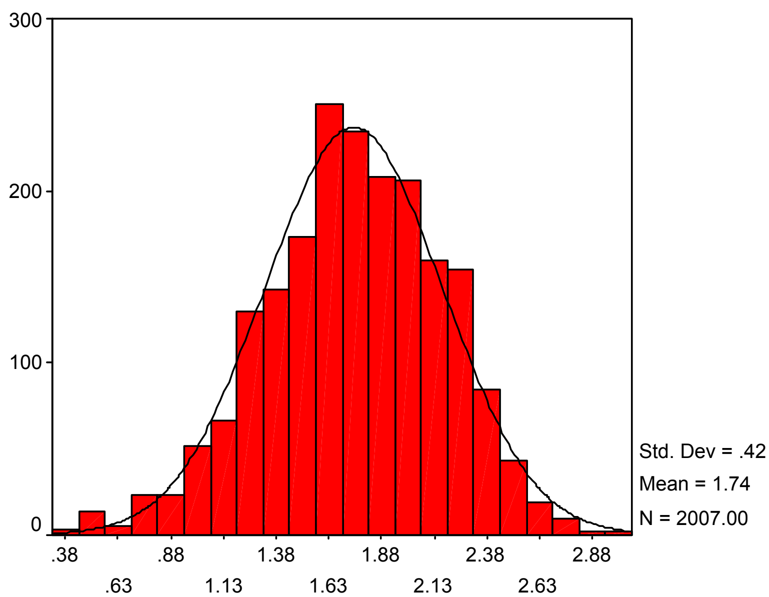
Modeling Auto Ownership

The results of the model indicate that the largest cost for most households in this country is the cost of owning a car. This is the main factor for how much a household drives. We start with auto ownership from Census2000 as the dependent variable, and look at how this correlates to our independent variables, and develop a formula to estimate how many autos the average household in a census block-group will own.

Auto ownership:

In the Minneapolis/St Paul MSA we find that the average Household auto ownership is 1.74 cars per household. The following histogram shows the distribution of auto ownership by census block-groups, according to the U.S. Census 2000:

Figure 11. Vehicles per Household by Census Block Group in Minneapolis-St. Paul



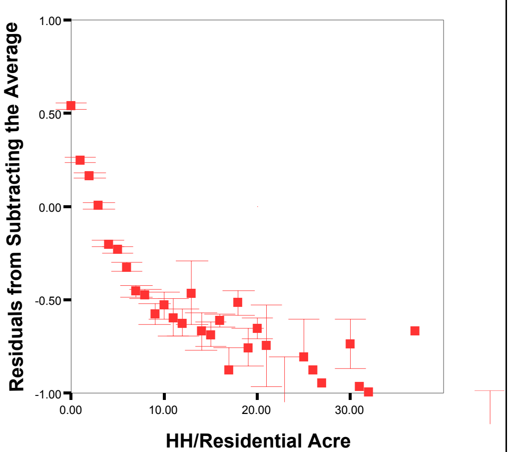
The following Table provides the descriptive statistics for Autos per Household in the Minneapolis/St. Paul Metropolitan Statistical Area (MSA).

Table 3: Auto Ownership Descriptive Statistics

Vehicles / Household			
Block Groups	Mean	Minimum	Maximum
2007	1.74	.35	3.00

The previous chart and table showed the actual distribution, as measured by the Census, for Auto Ownership in Minneapolis. The following seven graphs show the relationship of each of Auto Ownership with each of the other local environment variables, which will inform the structure of the model. The graphs are “Error Bar Graphs”, which show the average of the y value at the center of each point, with the average standard error shown by the line intersecting each point. Note that each point represents a bin of data whose center is the value of the x-coordinate. The variation seen in each graph is the objective of the regression analysis. The functional form of this variation is obtained by examining what reasonable function will fit the distributions shown.

Table 4. Relationship of Auto Ownership with the other Local Environment Variables



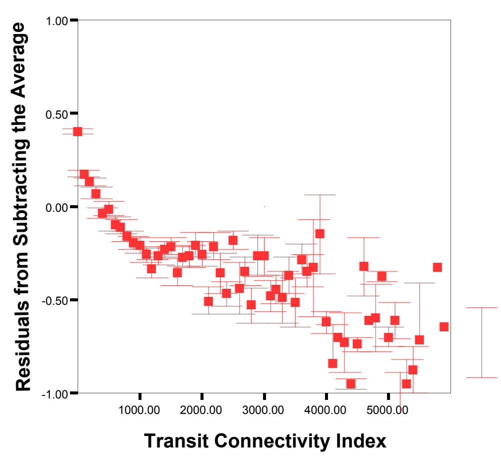
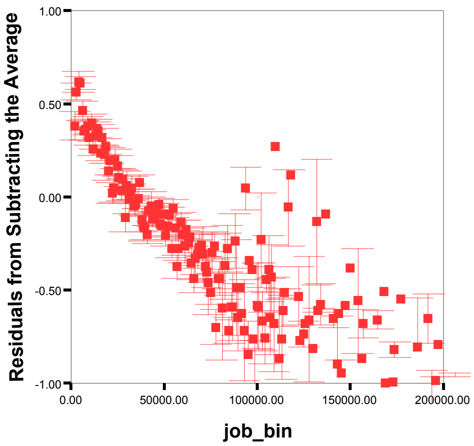
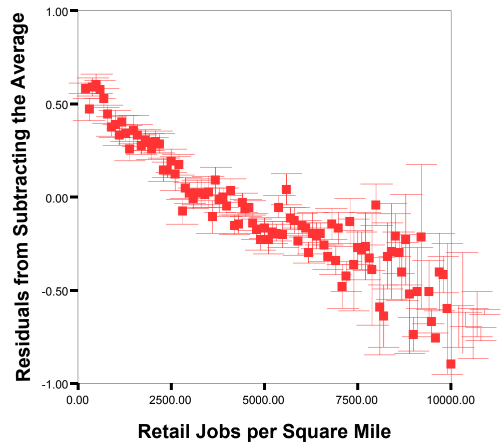
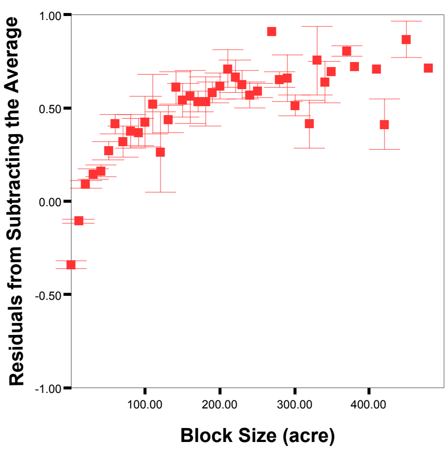


Table 4. Relationship of Auto Ownership with the other Local Environment Variables







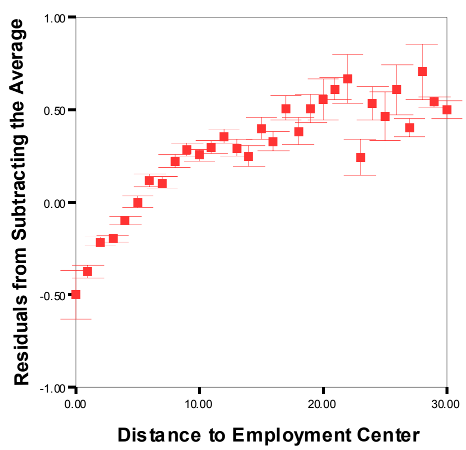
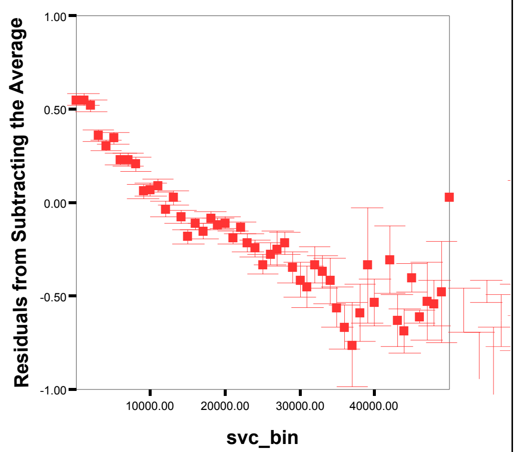


Table 4. Relationship of Auto Ownership with the other Local Environment Variables



The next table shows the chosen formula describing the dependencies observed for each of the seven independent local environment variables:

Table 5: Dependencies of Local Environment Variables			
Variable	*Symbol*	*Formula*	*R^2*
Average Block Size	BlkSize	n + s (BlkSize t)	36%
Distance to Nearest Large Employment Center	Rec	n + (s*ln(Rec+t))	37%
Number of Retail Jobs per Square Mile	JRetail	n + s*(t** JRetail)	44%
Number of Service Sector Jobs per Square Mile	JSvc	n + s*ln(JSvc + t)	50%
Number of all Jobs per Square Mile	JTotal	n + s*ln(JTotal + t)	52%
Transit Connectivity Index	TCI	n + (s*ln(TCI+t))	53%
Residential Density	DR	n + (s*ln(DR+t))	61%

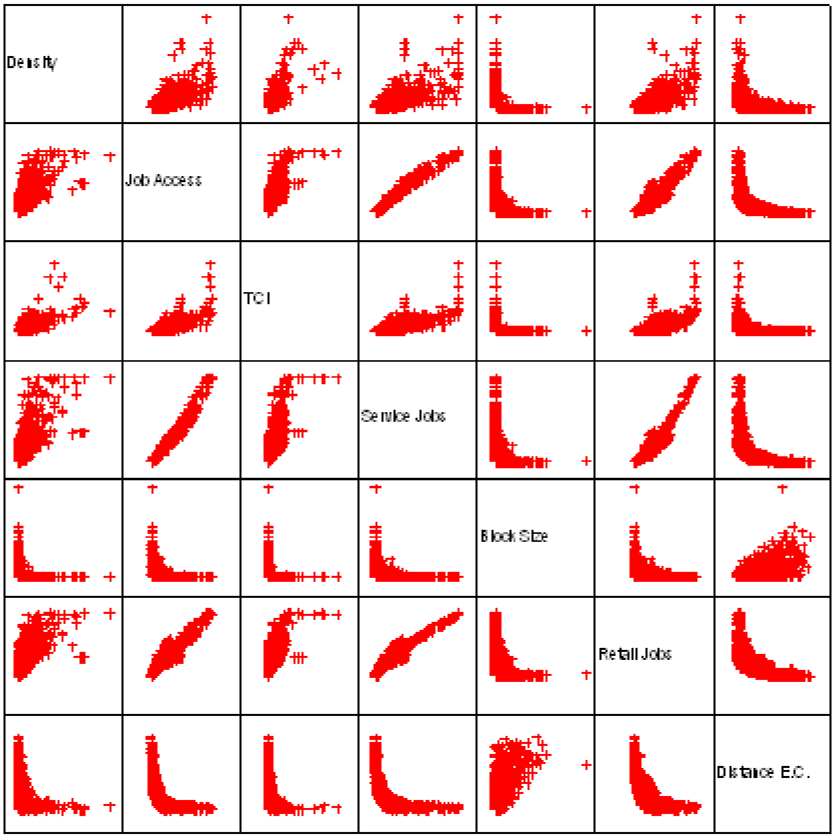
In order to find a combined formula that will describe auto ownership as a function of these parameters we use a complete multi-dimensional regression analysis. The combinations of the above formulas that improve the R^2 for the fit are shown below in Table 6.

Table 6: Combined Regression Analysis and Formula for Auto Use		
Variables in fit	*Formula*	*R^2*
Residential Density	$N + S * \ln(D_R + T)$	61%
Residential Density, and Number of all Jobs per Square Mile	$N + S * \ln(D_R + T) + (U*\ln(J_{Total}))$	64%
Residential Density, Number of all Jobs per Square Mile, and Transit Connectivity Index	$N + S * \ln(D_R + T) + (U*\ln(J_{Total})) + (W * \ln(TCI + X))$	65%
Residential Density, Number of all Jobs per Square Mile, Transit Connectivity Index and Average Block Size	$N + S * \ln(D_R + T) + (U*\ln(J_{Total})) + (W * \ln(TCI + X)) + ((Blk_{Size}{}^{Z}))$	66%
Residential Density, Number of all Jobs per Square Mile, Transit Connectivity Index, Average Block Size and Distance to Nearest Large Employment Center	$N + S * \ln(D_R + T) + (U*\ln(J_{Total})) + (W * \ln(TCI + X)) + ((Blk_{Size}{}^{Z})) + (A*\ln(R_{ec}))$	66%

$$N + S * \ln(D_R + T) + (U*\ln(J_{Total})) + (W * \ln(TCI + X)) + ((Blk_{Size}{}^{Z})) + (A*\ln(R_{ec}))$$

The last formula in the table, and repeated above is the most effective combination of variables of formulae that are significant in reducing the R^2. The reduction in R^2 does not reduce with as big an effect after adding each variable as it may appear in the above table. This is because all of these variables are highly correlated, and reducing the variation in auto ownership by one variable will take away some correlation with another variable. The following two figures show high correlation among variables.

Figure 12. Correlation between each of the local environment variables



Since most of the variables are highly correlated with one another, once one is fit each additional variable does not show as big a dependence. As an example, the following seven plots show the results of only fitting to residential density:

Table 7. Auto Ownership Fit with Residential Density

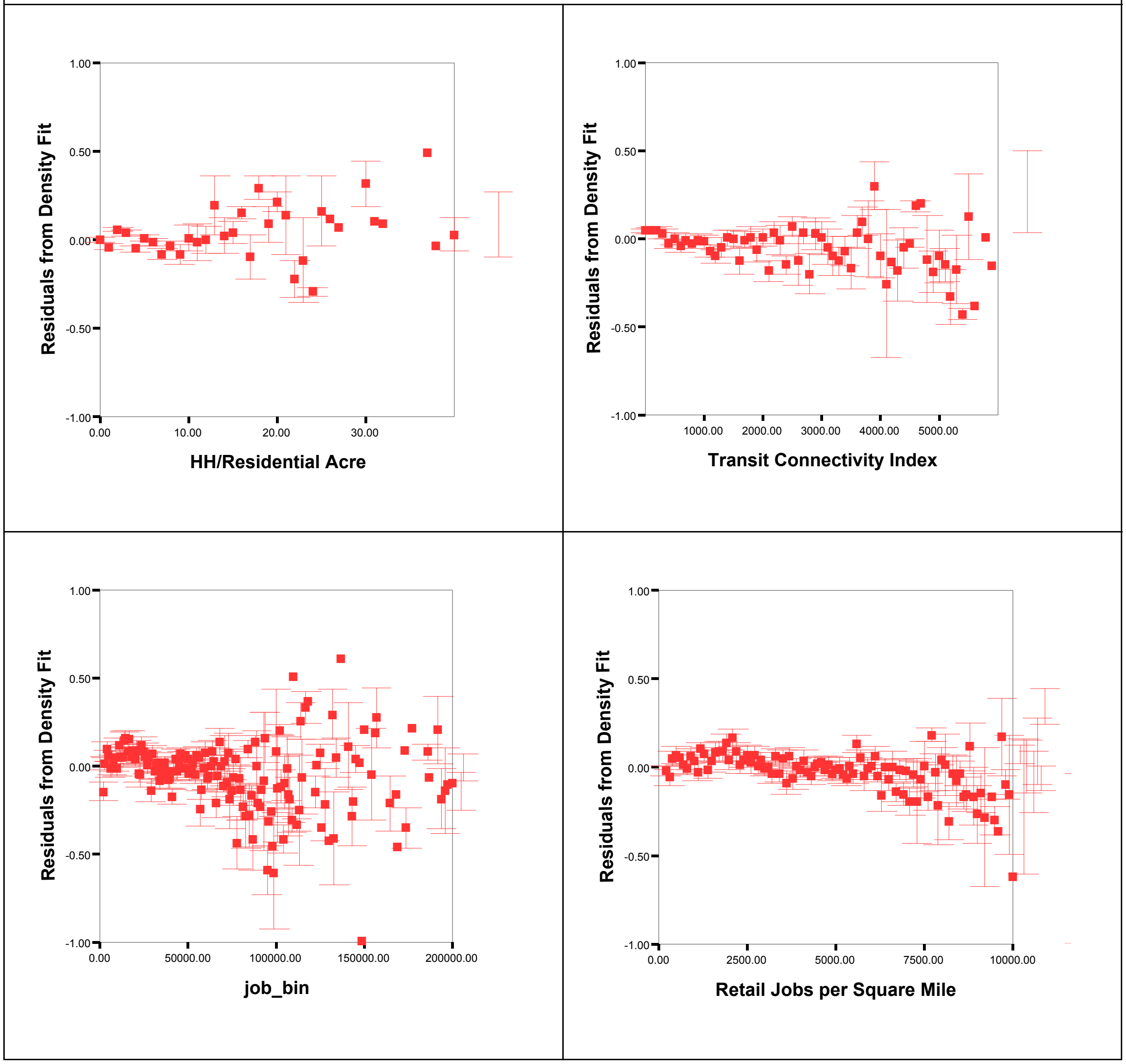
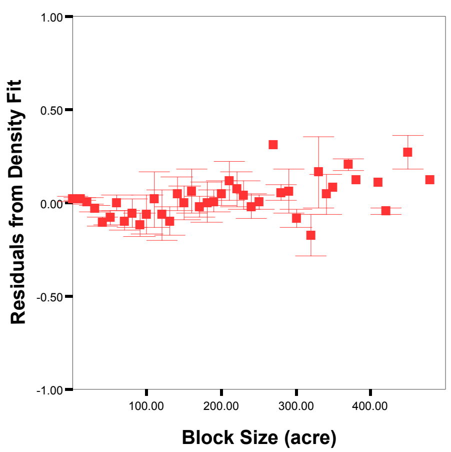
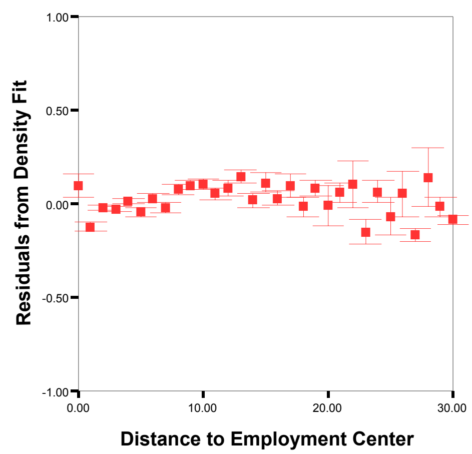
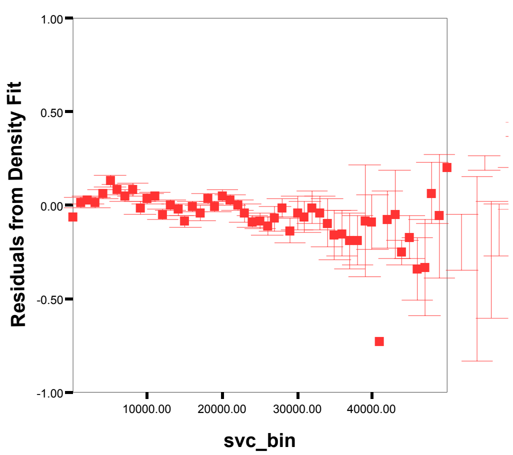


Table 7. Auto Ownership Fit with Residential Density







Using the above regressions, we determined the following formula:

$$F(V_{uf}) = N + S * \ln(D_R + T) + (U*\ln(J_{Total})) + (W * \ln(TCI + X)) + ((Blk_{Size}^{Z})) + (A*\ln(R_{ec}))$$

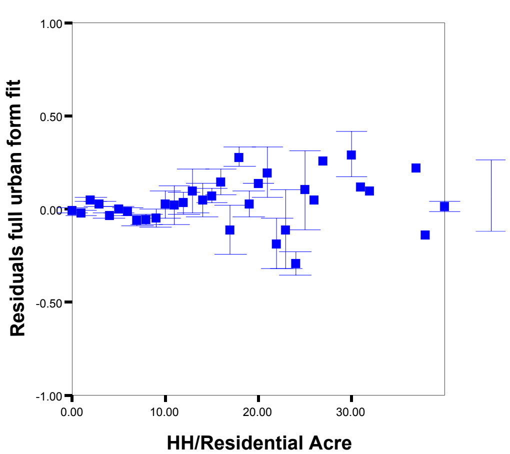
optimized the model, where the fit parameters N, S, T, U, W, X, Z and A are optimized to minimize the R^2 at the values in the following table.

Table 8: Auto Ownership Fit Parameter Values for Local Environment Variables		
Parameter	*Value*	*Error*
N	2.70	.25
S	-.350	.021
T	.507	.10
U	-.0484	.018
W	-.1163	.021
X	237.	114.
Z	-.161	.029
A	.0218	.013
R^2	**65.7%**	

The resulting **R^2** value, 65.7% represents approximately 2/3 of the variation in household auto ownership is explained by the local environment variables. The following seven graphs show that the formula expressed above does indeed flatten out the distributions above.

Much like the previous graphs these are "Error Bar Plots." However, the y-axis is the residual from the regression analysis – the difference between the fit value and the measured value. If the formula is correct these graphs will look flat and be centered on the y-axis coordinate zero. Note: while this is true for all of these graphs - the variation remaining is either statistical variations or remaining dependencies of variables for which we have corrected.

Table 9: Residual of local environment variable fit for each of the seven local environment variables



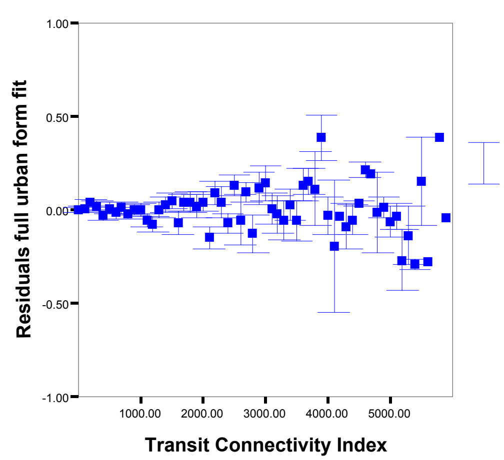


Table 9: Residual of local environment variable fit for each of the seven local environment variables

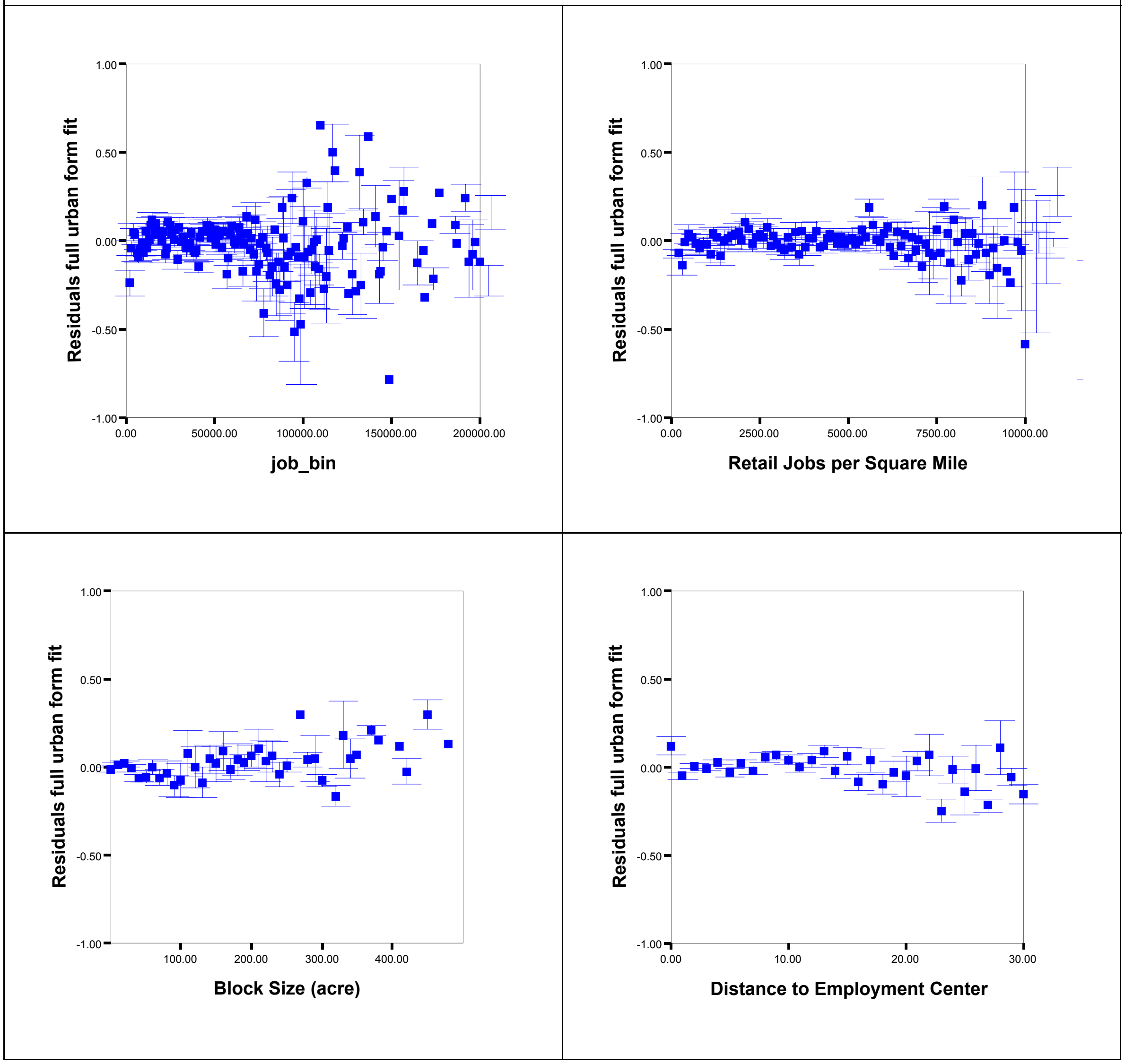
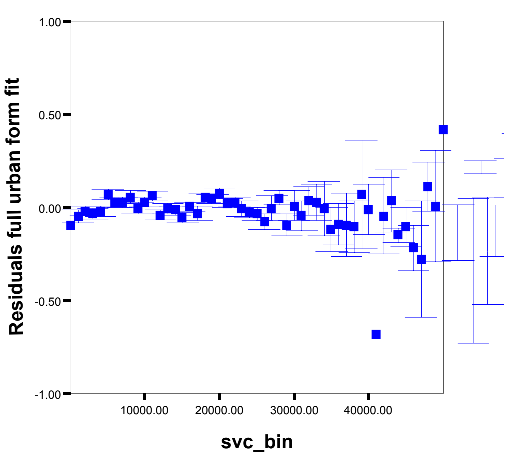
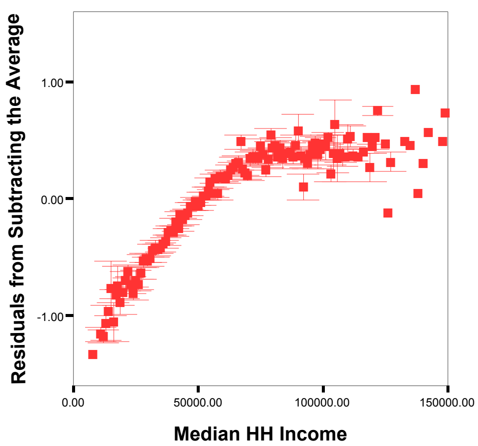


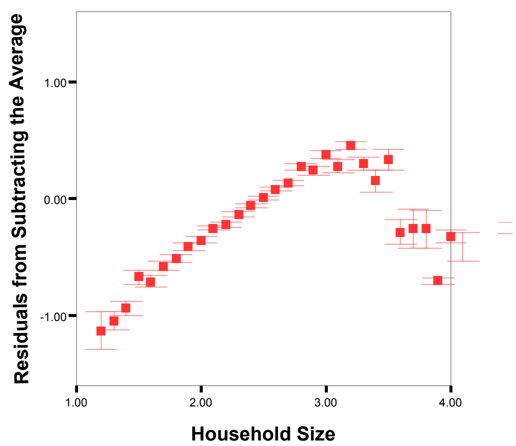
Table 9: Residual of local environment variable fit for each of the seven local environment variables	
	

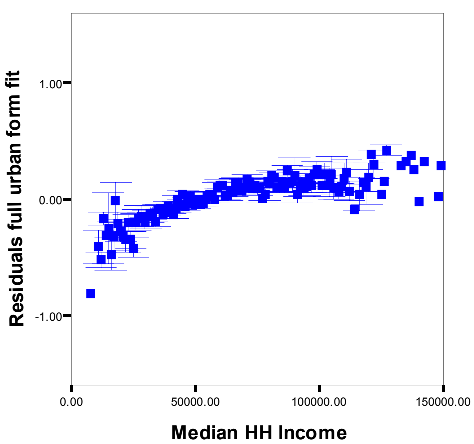
We have reduced the variability of auto ownership caused by the local environment, but it is also important to incorporate the variation caused by the *household variables*; household income and household size.

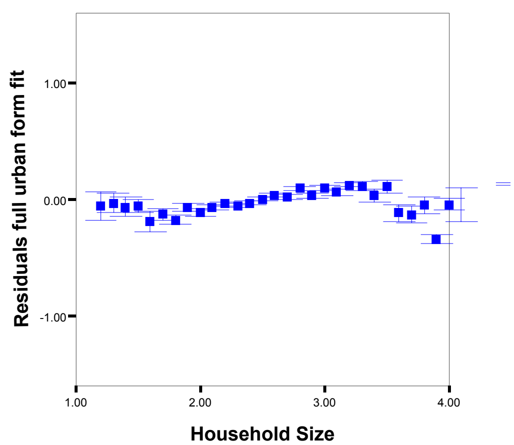
The following two plots show that fitting for the local environment variables reduces the variation in these household variables. The red graphs in row one show that auto ownership varies by household income and size, and the blue graphs (row two) show how this variation is reduced once the local environment variation has been removed. ***This is a remarkable finding that most of the variability is removed by the local environment variable and not the individual households, showing just how much place matters!***

Table 10: Variation in Auto Ownership with Household and Local Environment Variables









However, there is still some dependence on mostly household income and size and the following functional form will indeed decrease it:

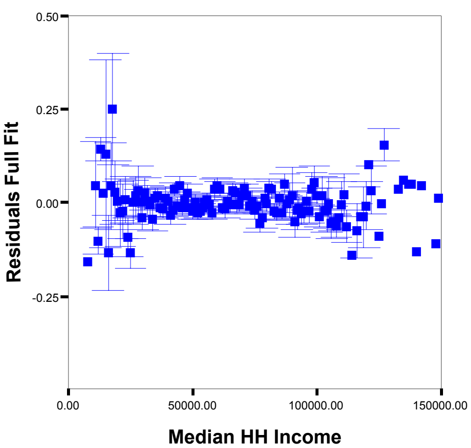
$$G(V_{hh}) = N+\exp(a+(b/Income_{hh}))+\exp(d/Size_{hh})$$

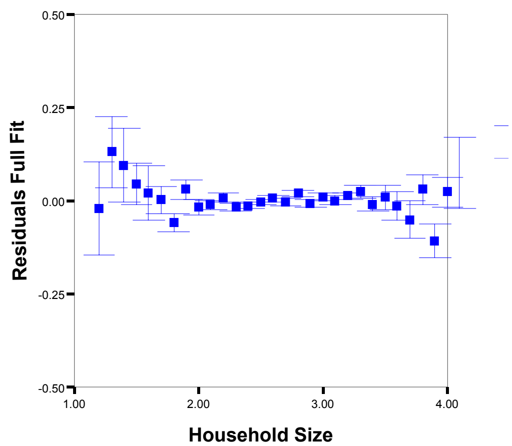
Where the fit value for the parameters N, A, B and D are:

Table 11: Fit Values for Variation in Auto Ownership with Household Characteristics		
Parameter	*Value*	*Error*
N	-.809	.238
A	-.0106	.231
B	-10,107.5	3,378.0
D	.00890	.0384
R2	24%	

The following graphs show the result of this fit of auto ownership with household variables:

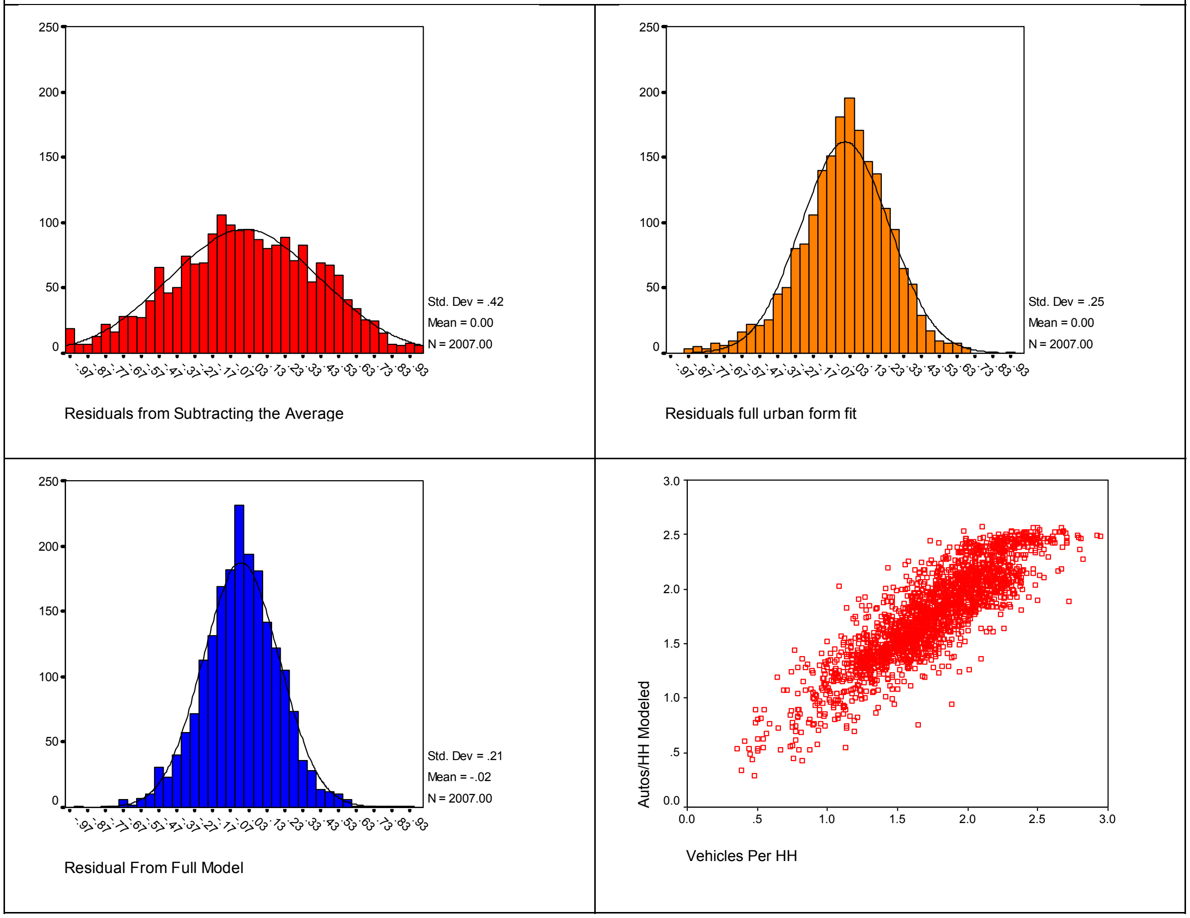
Table 12: Residuals of Household Variable Fit with Auto Ownership





Finally, the following graphs show the overall reduction in variation with this calibrated model for the Auto Ownership component of transportation costs. The first three graphs show that with more refined fitting the overall variation is reduced and the last shows the predicted vs. measured auto ownership

Table 13: Reduction in Variation for the Auto Ownership Component of the Model



Modeling Transit Use:

The next component of household transportation cost is transit use. Since there is no direct measure of how many transit rides are purchased everyday we used the journey to work data that is gathered by the US Census and data on system ridership from the 2001 FTA Database[8]. The method of determining transit rides per household per day is:

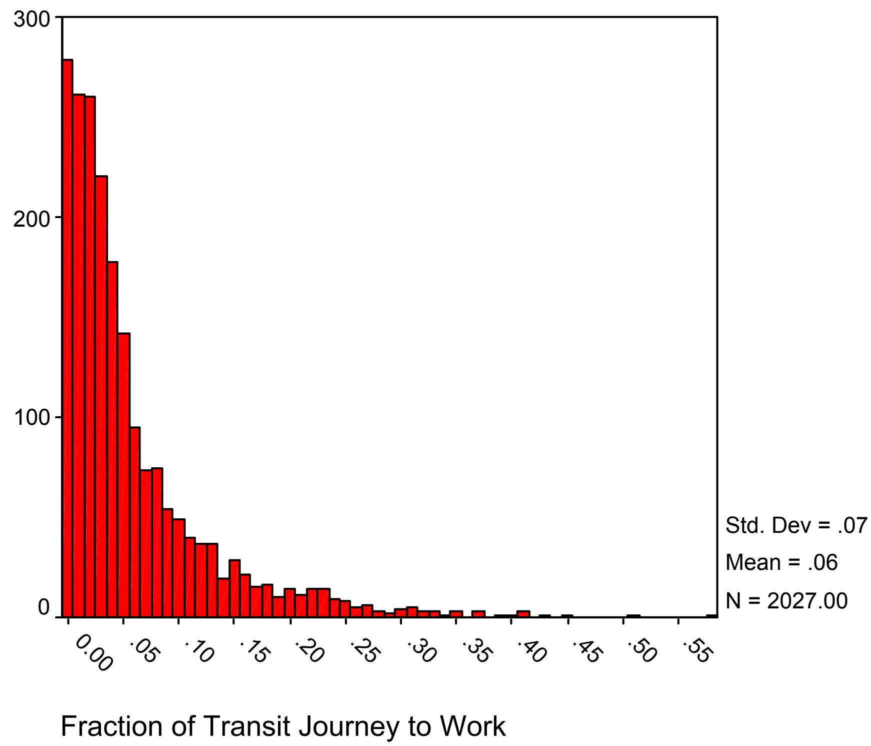
- Determine fraction of "Journey to Work" trips (where fraction of journey to work is f_{jtw}) that use transit by census block groups and calibrate a model that will predict this by Census block groups
- Look at the total number of transit trips in the region from the FTA database and then scale up the journey to work transit numbers per Census block group so that the total trips for all the households summed over all the Census block groups add up to the overall transit trip number.

[8] http://www.ntdprogram.com/ntd/Profiles.nsf/2001+All/5027/$File/P5027.htm

Fraction of Journey to Work Trips using Transit

Initially it would seem that to calibrate a model that will predict f_{jtw} would be very similar to what was just accomplished for the auto ownership model. However it is a little more complicated because this particular statistic follows either the binomial or Poisson distribution, not the normal distribution that was so clear from the plots for Auto Ownership. The following histogram show that this distribution is peaked at zero, and for obvious reasons it cannot be less than zero (one cannot take a negative transit trip – even though some will claim to have done so!).

Figure 13: Distribution of Journey to Work Trips in Minneapolis-St. Paul



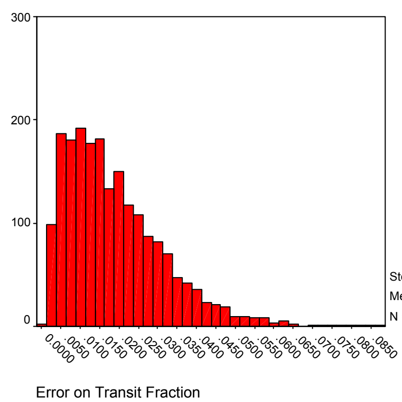
If one is to fit this distribution using a simple regression method, the modeled result will be biased in such a way to always reduce the large number and increase the low numbers. The way we have dealt with this is to use the error on each measurement a weight for the loss function in the regression. To determine the error for f_{jtw} is straight forward considering that this is a binomial statistic (true for transit trips, false otherwise). Note that this ratio is obtained from the Census-2000 SF 3 long form. The following formula is suggested by the Census Bureau to calculate the error on a ratio of this nature given the sampling rate of approximately 1 to 6 is:

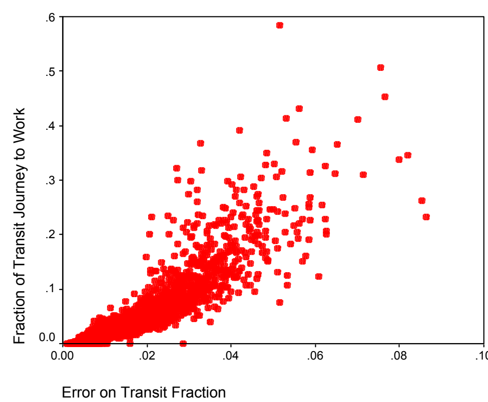
> "For a percentage and/or base of percentage …, use the formula given below to calculate the standard error…
>
> $$SE(P) = Sqrt((5/B)*P*(100 – P))$$
>
> B = Base of estimated percentage
>
> P = Estimated percentage
>
> The 5 in the above equation is based on a 1-in-6 sample and is derived from the inverse of the sampling rate minus
>
> one, i.e., 5 = 6 - 1."[9]

One outstanding problem is for block groups where f_{jtw} is zero – what is the standard error – it cannot be zero, since there is some chance that they have not yet sampled enough to measure a positive event (a transit trip). The solution for this is to assume, for reasons of calculating the standard error only, that there is one transit trip and use that ratio for P above (in other words P = 1/B).

The histogram and scatter plot below depict the standard error of the fraction of transit trips. These graphs show that the error on the measurement of the fraction varies up to 0.1 and that the larger values have a larger error, this is typical for binomial statistics.

Table 14. Standard Error of Transit Trips

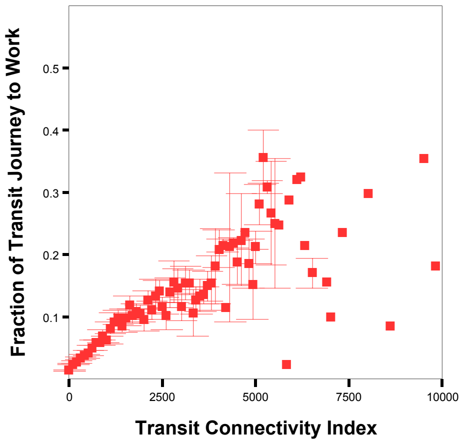


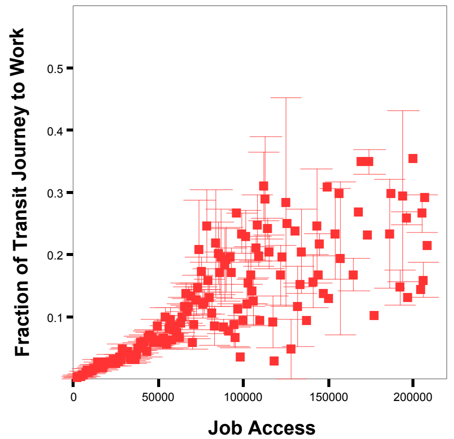


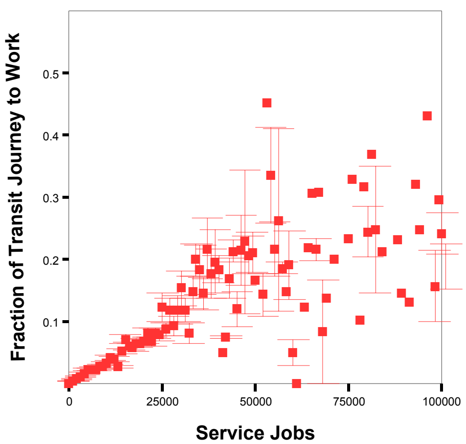
The plots below show the relationship of f_{jtw} with the seven local environment variables.

[9] http://www.census.gov/prod/cen2000/doc/sf3chap8.pdf

Table 15: Variation in the Journey to Work with the other local environment variables







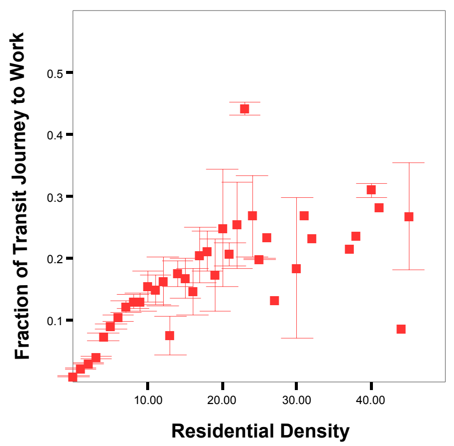
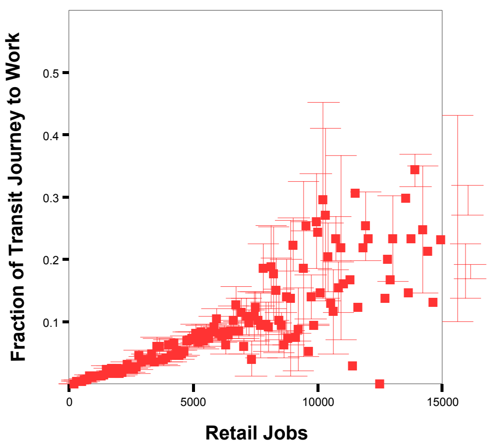
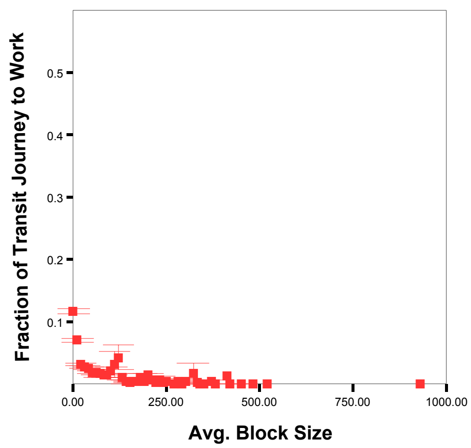
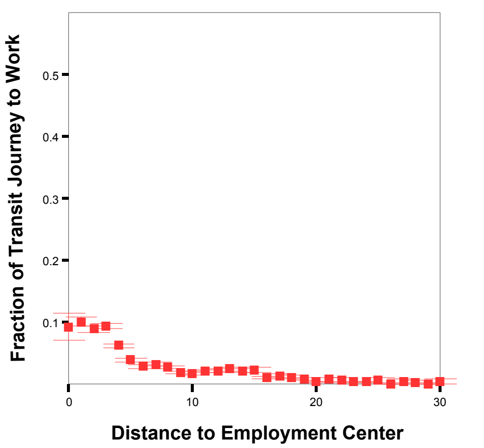


Table 15: Variation in the Journey to Work with the other local environment variables







The local environment variables can be fitted to Transit Use with the following functions:

Table 16: Formulas to Fit Local Environment Variables to Transit Use

Variable	*Symbol*	*Formula*	*R^2*
Average Block Size	Blk_{Size}	n + s*e** (Blk_{Size}+t)	30%
Distance to Nearest Large Employment Center	R_{ec}	n + (s*e**(R_{ec}+t))	24%
Number of Retail Jobs per Square Mile	J_{Retail}	n + s*(t** J_{Retail})	38%
Number of Service Sector Jobs per Square Mile	J_{Svc}	n + s*e**(t/J_{Svc})	47%
Number of all Jobs per Square Mile	J_{Total}	n + s*e**(t/J_{Total})	50%
Transit Connectivity Index	TCI	n + (s/(TCI+t))	52%
Residential Density	D_R	n + (s/(D_R+t))	45%

Like for the auto ownership fits the fits to these variables increase the likelihood of the fit, but with diminishing returns. The final formulation is:

$$f_{jtw} = n + (a/(b+ TCI)) + (\exp(c + (d/ J_{Total}))) + (e/(f+ D_R)).$$

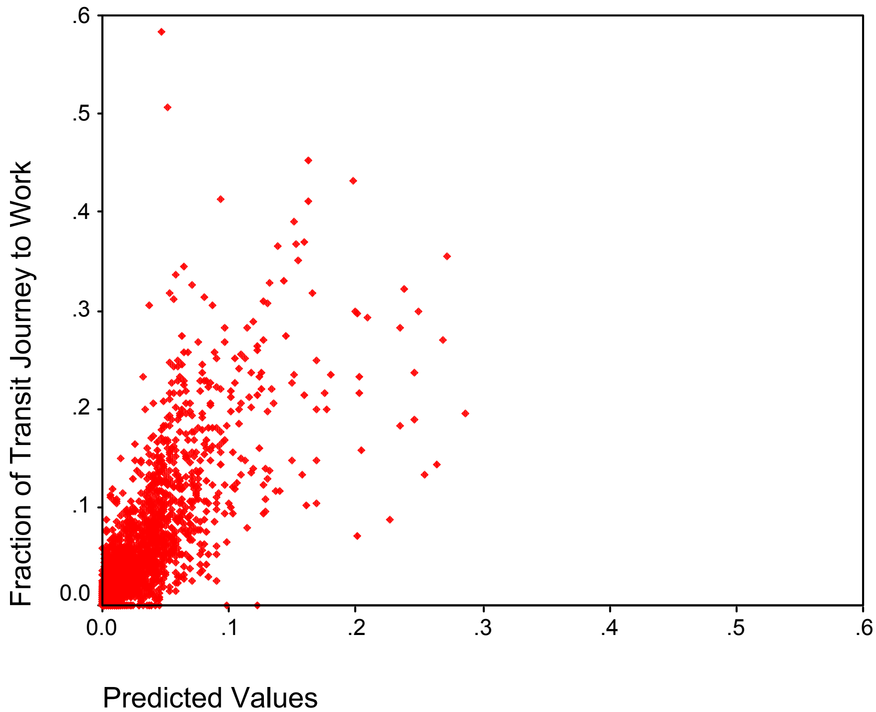
As mentioned above we used a different loss function for the regression analysis:

$$(R/SE)^2$$

This method, known as “Chi Square Fitting,”[10] will give a less biased result. This is illustrated in the following plots.

[10] See for example - NUMERICAL RECIPES IN C: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43108-5) Copyright (C) 1988-1992 by Cambridge University Press. Programs Copyright (C) 1988-1992

Figure 14. Predicted Values of Transit as a Fraction of Journey to Work

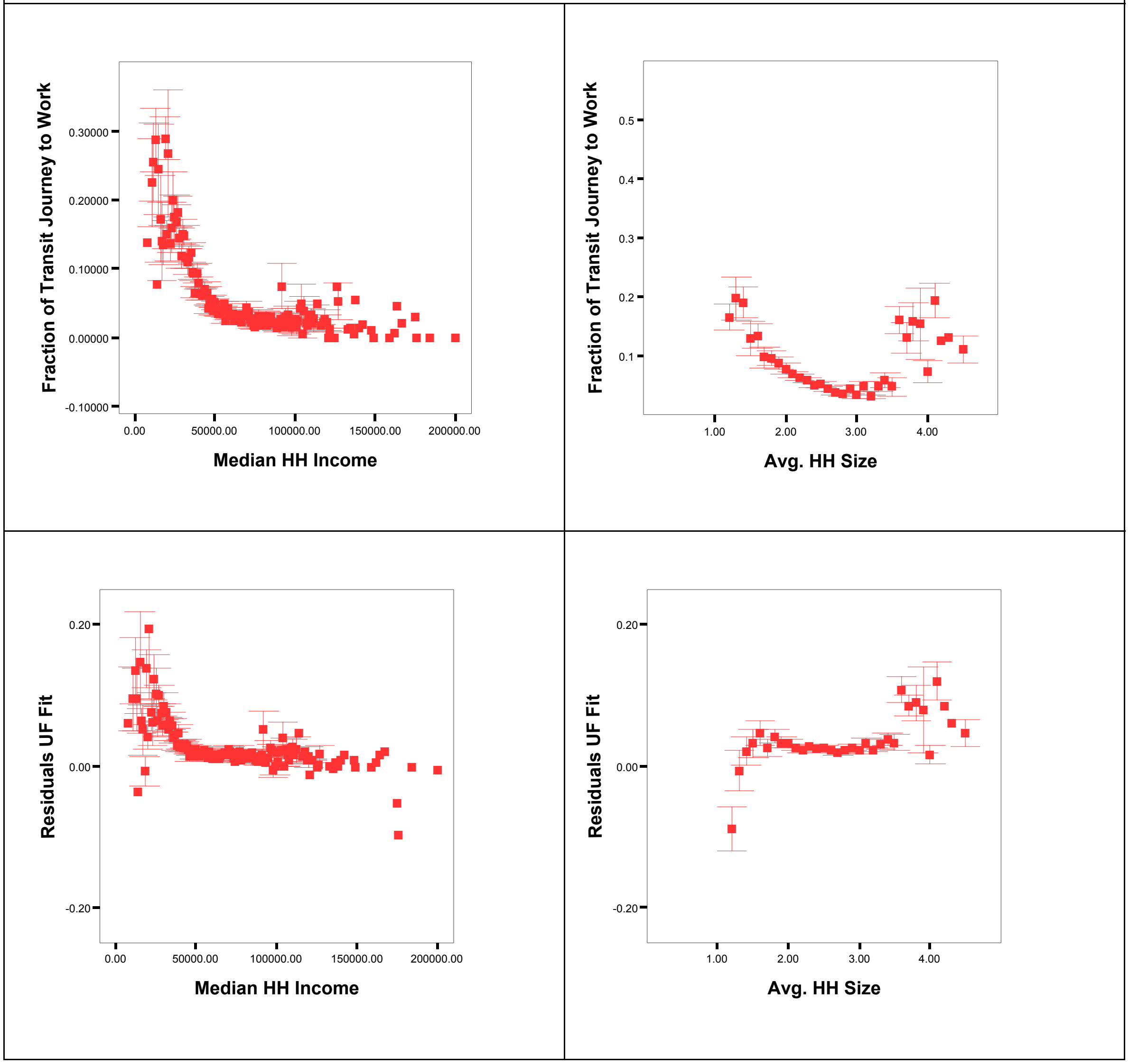


The final fit values are shown in the table below:

Table 17: Transit Fit Parameters for Local Environment Variables	
Parameter	*Value*
N	.71
A	-1800
B	10354
C	-1.84
D	-157064
E	-88
F	164
Chi Square/Degree of Freedom	7763/(2006 – 7) = 3.9

The next series of plots depict the relationship of the fraction of journey to work trips by transit (f_{jtw}) in comparison to the household variables:

Table 18: Fraction of Journey to Work Trips by Transit Compared to Household Characteristics



Note that again most of the variation of the f_{jtw} is being found from its relationship with the local environment variables. However, the remaining variation must be calibrated for this model. Since the value of this fraction if often zero, we need to vary this only slightly with the Household variables. The best way to combine the local environment function with the household function is a sum rather than a product – in other words:

$$PT = F(V_{uf}) + G(V_{hh})$$

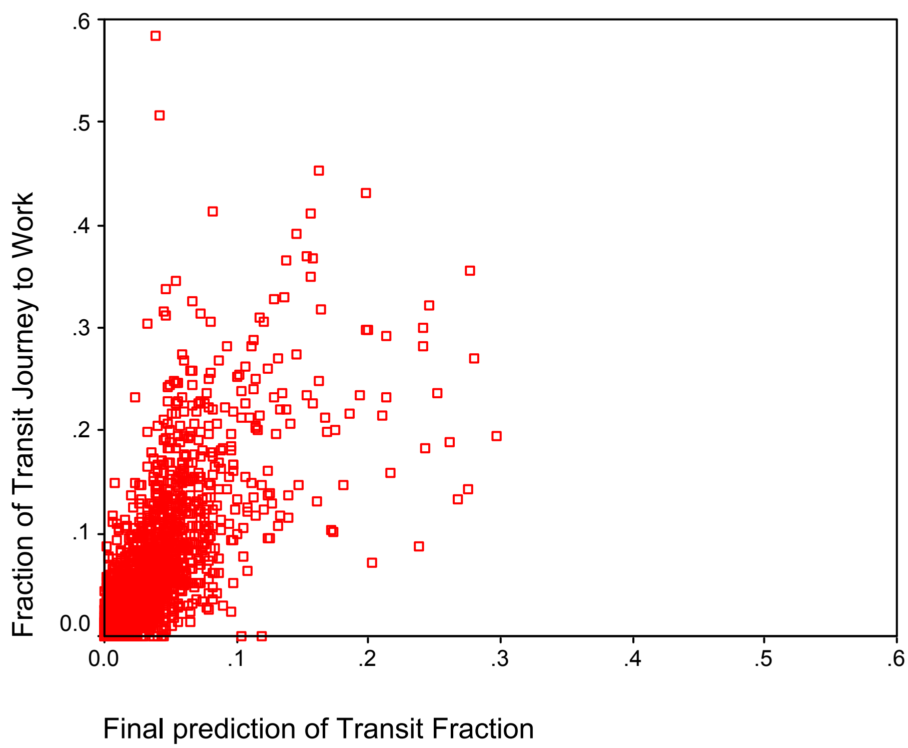
where the functional form for G is: $G(V_{hh}) = N + (a / (b + Income_{hh})) + c^* \ln(d + Size_{hh})$

and N, A, B, C and D are the values of the fit parameters, shown below:

Table 19: Transit Fit Parameters for Household Variables	
Parameter	*Value*
N	.013
A	491
B	14002
C	-.065
D	.54
Chi Square./Degree of Freedom	7640/(2006-5)= 3.8

The final plot of predicted fraction of journey to work trips by Transit versus the measured values in the Census 2000:

Figure 15. Plot of predicted versus measured for the fraction of journey to work trips by transit



Calculating the transit rides per day per household:

In order to translate the fraction of journey to work trips by transit into a total transit use per household, we use the FTA's[11] transit data base for 2001, which supplies several counts of total transit trips for Metro Transit. These will then be compared to the population using transit for work in order to estimate total daily transit trips per household. Table 20 displays the data available from the FTA for the Minneapolis-St. Paul Transit System, and our interpretation of total ridership into rides per household.

Table 20: Total Transit Ridership Statistics for Metro Transit in 2001

Service Consumption	
Annual Passenger Miles	312,516,373
Annual Unlinked Trips	73,347,859
Average Weekday Unlinked Trips	243,998
Average Saturday Unlinked Trips	129,226
Average Sunday Unlinked Trips	81,947
Services Supplied	
Annual Vehicle Revenue Miles	25,147,036
Annual Vehicle Revenue Hours	1,839,659
Vehicles Available for Maximum Service	953
Vehicles Operated in Maximum Service	792
Base Period Requirement	308
Calculated Quantity	
Average passenger per revenue mile	12.43
Average annual passenger trips	911,535,135
Average annual passenger trip per household	893.31
Average daily passenger one-way trips per household	2.45
Calculated From these fits	
Sum of transit journey to work fraction * households	31970.4
α	2313

To assign the average daily number of transit trips per household by census block group the assumption that this is proportional to the f_{jtw}, or

[11] http://www.ntdprogram.com/ntd/Profiles.nsf/2001+All/5027/$File/P5027.htm

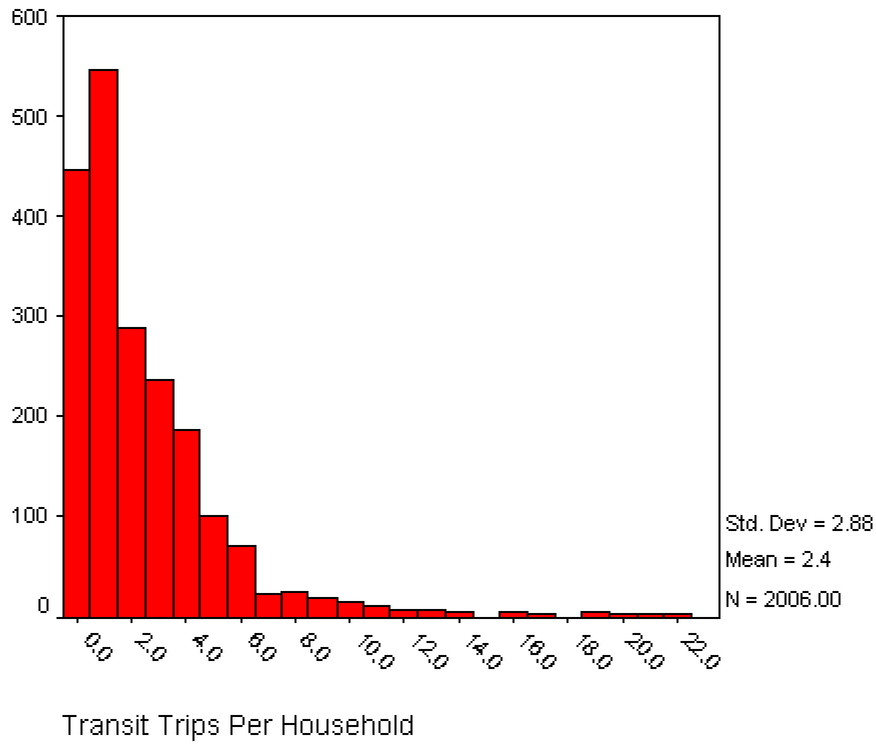
$$\text{Trips/HH} = \alpha * F_{jtw}$$

Where α is a proportionality factor. Note that:

$$\alpha = \text{Total regional Transit Trips/ Sum}(f * h)$$

This is calculated above using the modeled value for f from the fit. Once this is calculated the distribution of Transit trips per household looks like:

Figure 16. Distribution of Transit Trips per Household



Modeling Auto Use – Vehicle Miles Traveled (VMT)

"How much do you drive your car?" Before we could model auto use, we needed a valid source for total vehicle miles traveled per household. Unfortunately, there is no clear data source with the definitive answer at the right geographic scale. In previous studies we had access to odometer readings by zip+4, which allowed us to accurately measure the average vehicle miles traveled in a small area, such as a census tract or a block group. Our first efforts were focused on getting this kind of data for the Minneapolis/St Paul MSA, but we discovered that even though those data had been collected from the auto inspection program, they had not been saved when the program was discontinued several years ago. Given this, we set out in search of a data source that we could use for the dependant variable that represents Auto Use. We do have two years of odometer readings this from the Chicago metro area, but decided we needed a more ubiquitous source since this model will be used nationally. The US-EPA published the total miles driven by county for the entire US, but it is not the miles driven by autos owned in each county but rather the autos that were driving through a county. This led to large amounts of VMT in places like South Dakota, where there are not many people or cars, but there is an interstate; and in very urban areas, like Atlanta or Chicago, where household auto ownership is low but many interstates converge and there is high job density.

The 2001 National Household Transportation Survey (NHTS)[12] provided a solution. The data set (available for download) provides very detailed information on the amount of household travel information, including autos owned and miles driven. However, it is not reported to a small geography where we could get an appropriate set of independent variables from which to model VMT, but there are seven distinct fields in the database about each block group that we were able to match to all Census block groups. Through this matching process, we identified the actual block group of a significant number of surveyed households. This matching method was not intended by the NHTS publishers, as it breaks confidentiality of the survey respondents. Our use of the data, however, is not to identify the respondents, but to pair the local environment variables to the miles driven by certain types of households. The table below lists the seven fields and how we matched them to the Census.

[12] See: www.http://www.bts.gov/programs/national_household_travel_survey/

Table 21: NHTS Fields used to identify household VMT by block group					
Field	*Geography*	*NHTS's source*	*NHTS Range of values*	*Census source*	Description
State	State	Household location	1-79 Including Puerto Rico	State FIPS	*This field is the state FIPS code of the household*
Tract Pop/ Square Mile	Census Tract	Claritas[13]	50, 300, 750, 1500, 3000, 7000, 17000, 30000	Used the Census2000 Population divided by the square miles of land in each track	*Used the density and then assigned the appropriate value from the list of 8*
Tract HH/ Square Mile	Census Tract	Claritas	25, 150, 700, 2000, 4000, 6000	Used the Census2000 Households divided by the square miles of land in each track	*Used the density and then assigned the appropriate value from the list of 6*
Percent of renters in the Tract	Census Tract	Claritas	0%-100% to two decimal places	Used the fraction of renters form the tenure form Census2000	*Used the percentage rounded to two decimal places*
Block Group Pop/ Square Mile	Census Block Group	Claritas	50, 300, 750, 1500, 3000, 7000, 17000, 30000	Used the Census2000 Population divided by the square miles of land in each block group	*Used the density and then assigned the appropriate value from the list of 8*
Block Group HH/ Square Mile	Census Block Group	Claritas	25, 150, 700, 2000, 4000, 6000	Used the Census2000 Households divided by the square miles of land in each block group	*Used the density and then assigned the appropriate value from the list of 6*
Percent of Renters in the Block Group	*Census Block Group*	*Claritas*	*0%-100% to two decimal places*	*Used the Census2000 Households divided by the square miles of land in each track*	*Used the percentage rounded to two decimal places*

[13] As stated in the NHTS 2001 User Guide – Appendix Q (http://nhts.ornl.gov/2001/usersguide/index.shtml) "The data contained in these variables was derived from 2000 Census data and estimated forward to 2001 by Claritas, Inc. An annual demographic update is developed by this company to serve as a source of estimates of population, household, and housing unit characteristics. These estimates are made at relatively small units of geography, such as census tracts and block groups, which make this update effective for use in supplementing the NHTS data. The update is a comprehensive process that relies on a number of data sources, including regional and city planning agencies, federal agencies (e.g., Bureau of Labor Statistics, Bureau of Census, Bureau of Economic Analysis) U.S. Postal Service, the direct mail industry, the real estate industry, and experts in the fields of geographic information systems and mapmaking."

To assign each NHTS household to a census block group we matched the *State*, the *tract pop/sq-mile*, the *tract hh/sq-mile*, the *block group pop/sq-mile* and the *hh/sq-mile* listed for the NHTS with those variables in the Census, and then we demand that both the tract and the block group percentage of renters match within one percent to identify a positive match. Of the original NHTS sample of 69,817 households, 8,912 can be unambiguously placed in census block groups using this method. We then further narrowed this sample to just those where the household income, and the VMT is accurately reported, we also demand that the number of autos with good VMT data match the number of autos reported as owned by the household[14]. This reduced the sample size from 8,912 to 6,840. We then add up all of the VMT for the autos owned in the household and assign this as the Total VMT per Household that we use as the dependent variable in the following regression.

Regression Analysis of VMT with Local Environment and Household Variables

Using the 2001 NHTS Household VMT results explained above, we identified a relationship between driving and each of these five variables:

- Local Environment Variables:
  - Households per acre
  - Block size
- Household measures
  - Household size
  - Annual Household Income
- Auto ownership itself

The last variable – Auto Ownership – while being a household variable, is one that we have previously modeled, and therefore treat differently. Here our functional form will be better represented by:

$$T_{cost} = C_{tsp} * F(X_{le}) * G(X_{hh}) * H(X_{autos})$$

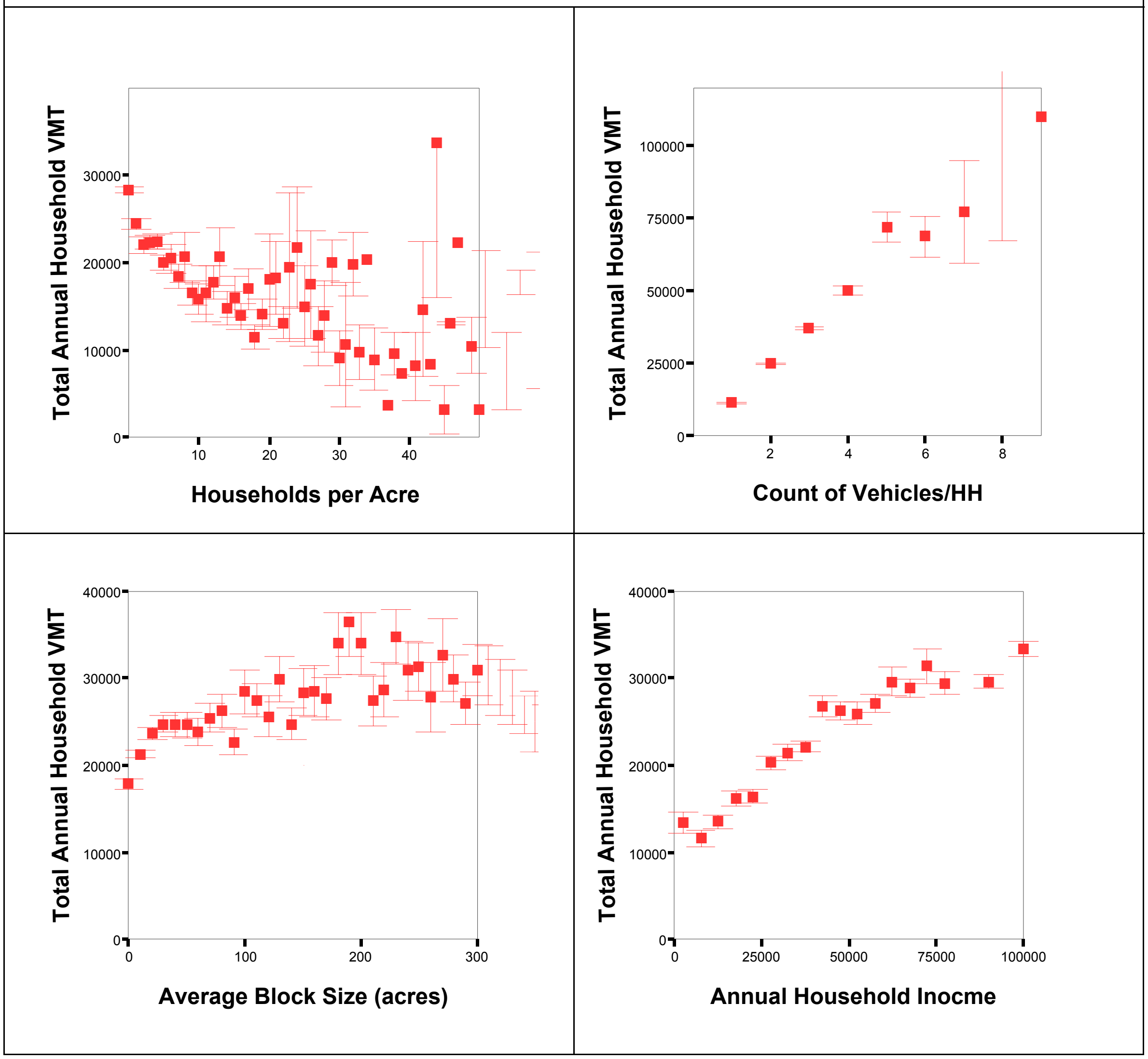
[14] See the NHTS 2001 User Guide – See Appendix B for definition of a valid income entry and Appendix J for a discussion of VMT, termed “BEST MILE” in the data.

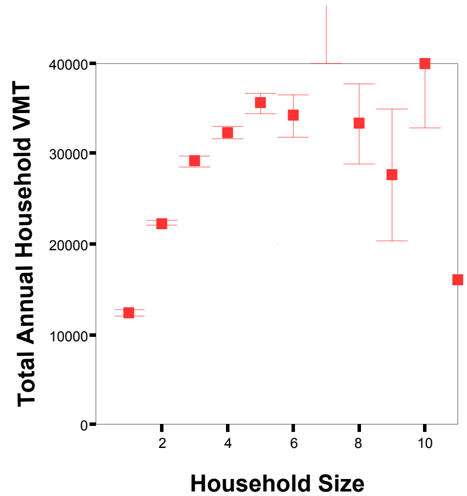
[15] See: www.http://www.bts.gov/programs/national_household_travel_survey/

Where the new Function $H(X_{autos})$ will be optimized and represent the component of driving variation related to the number of autos a household owns.

These five plots show the relationship of VMT to five local environment and household variables. These trends show driving clearly depends the most on the number of vehicles in the household.

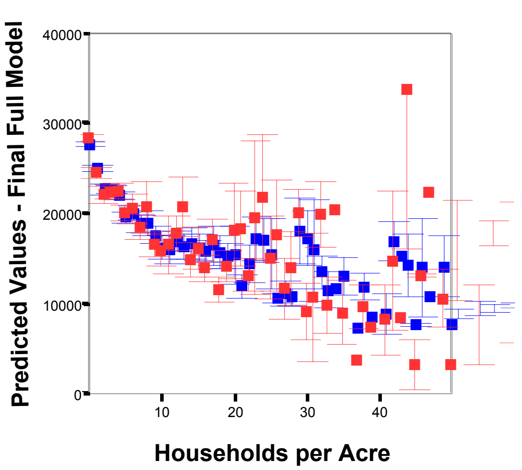
Table 22: Relationship of VMT to Five Local Environment and Household Characteristics





Just like the auto ownership model we fit F, G and H individually, and combine them in the end. It took two iterations for this component to obtain flat residual distributions. These next five graphs show (in blue) the modeled result, overlaying the actual data (in red).

Table 23: Modeled (blue) versus Measured (red) VMT per household



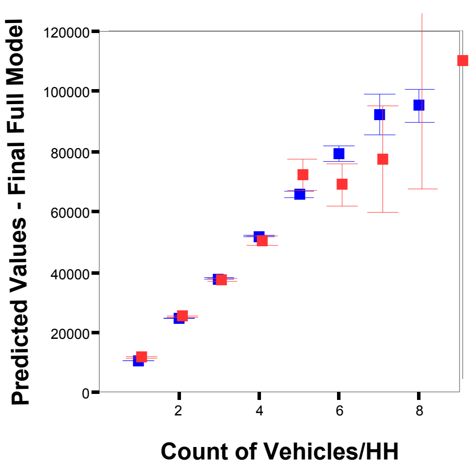
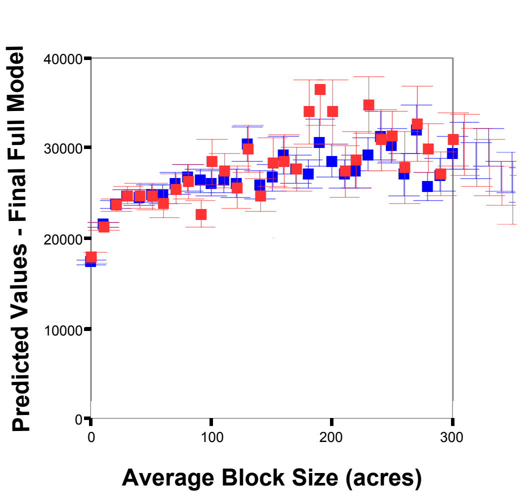
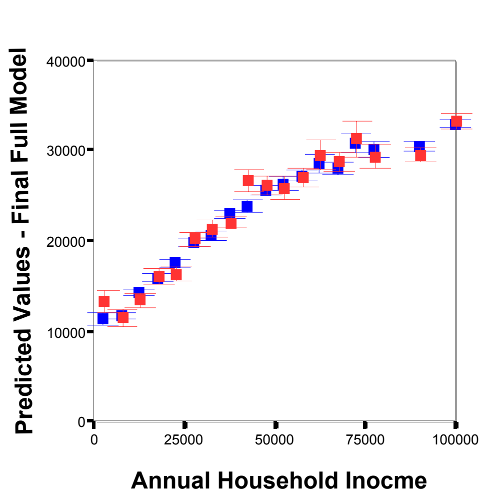
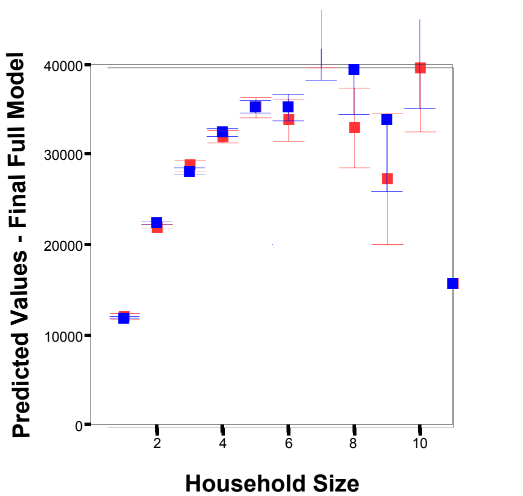


Table 23: Modeled (blue) versus Measured (red) VMT per household







Cleary this model predicts the trends with these variables. The following is the equations for the model:

Table 24: Formulas to Fit Local Environment Variables to Auto Use (VMT)		
Variable	*Equation*	*R^2*
Local Environment Variables		
Block Size	n * ln(ave_bloc)	0.036
HH/Raw Acre	n * ((hh_raw_a)**a)	0.035
Household Variables		
Income	n * ((hh_inc_e)**a)	0.092
Household Size	exp(a+ (b/ hhsize))	0.133
	Auto Ownership	
Autos/HH	n * ((hhvehcnt) **a) **this is h(#autos)**	0.341
Household and Local Environment Variables Combined		
	These next formula combine the above formulas:	
	prd_veh * (a+b*ln(ave_bloc)) * (1+c* ((hh_raw_a)**d)) **this is h*f(local environment)**	.34416
	prd_vuf * (b*((hh_inc_e)**c)) * (1 + d*exp(e/ hhsize)) **this is h*f*g(hh-variables)**	.37343

The complete equation for Auto Use is:

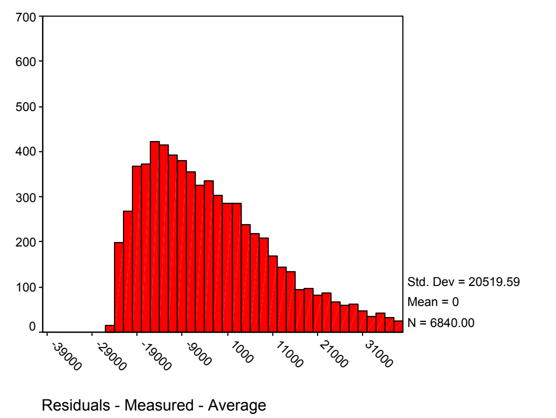
$$\text{VMT/HH} = n * (\text{Veh\_per\_hh}^{a}) *(b+c*\ln(\text{AvgBlock\_size})) *(1+d* (\text{hh\_per\_acre}^{e}))* (f*(\text{ave\_hh\_income}^{g})) *(1 + h*e^{(i/\text{ave\_hh\_sz})})$$

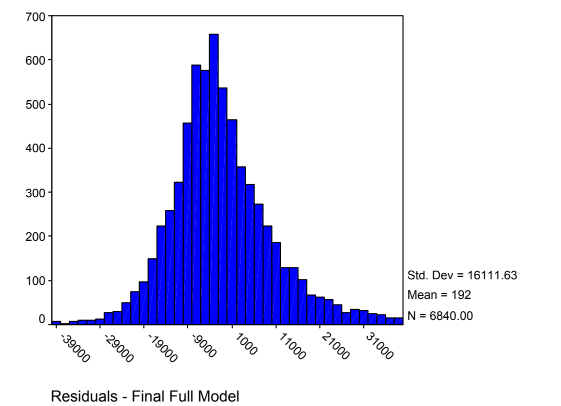
Where the following fit parameters for each of the variables are:

Table 25: Auto Use Fit Parameters for All Variables				
Veh_per_hh is vehicles per household	**AvgBlock_size** is the average size of the block in acres	**Hh_per_acre** is the households per land acre	**ave_hh_income** Is the average annula household income	**ave_hh_sz** is the average number of people in a household
And n, a, b, c, d, e, f, g, h, and i are all fit parameters whose values are:				
N = 13967	A = 0.893	B = 1.04	C = 0.0031	D = -0.0580
E = 0.1798	F = 0.6996	G = 0.0286	H = 0.1254	I = -3.0996

The following histograms show the data and the residual of the overall fit.

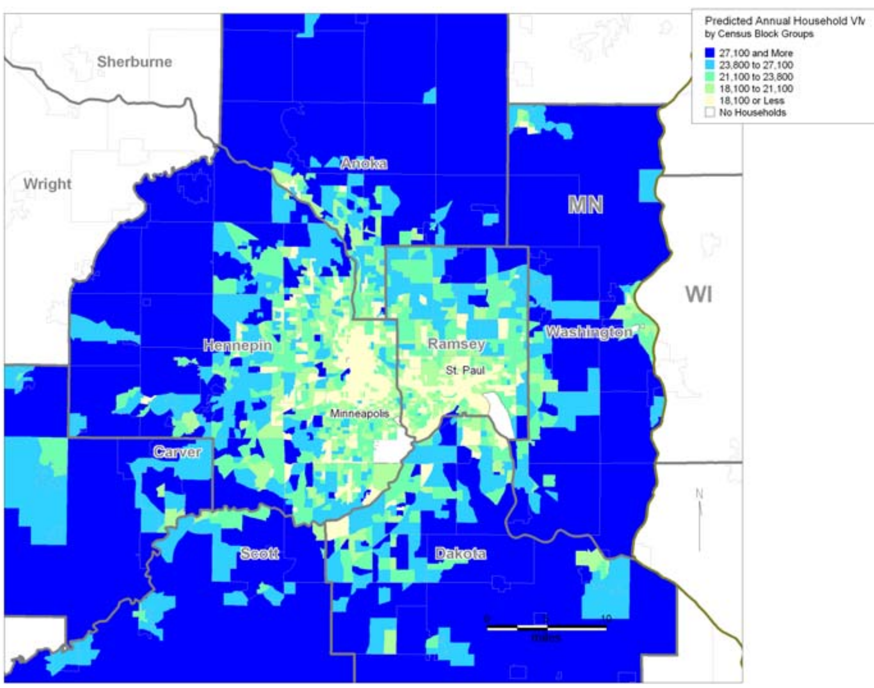
Table 26: Distribution and Residual of the VMT Fit





The following figure maps the model's prediction of total annual household vehicle miles traveled in the Minneapolis-St. Paul region.

Figure 17. Map of Total Household Annual VMT in Minneapolis-St. Paul



Model Results

Given that this model is so closely connected to a place, maps offer an ideal platform to summarize the results. The following section uses several maps to display the final results of the model—how the three transportation components summarize to total transportation costs and how they combine with housing costs. The index will ultimately address the following two questions:

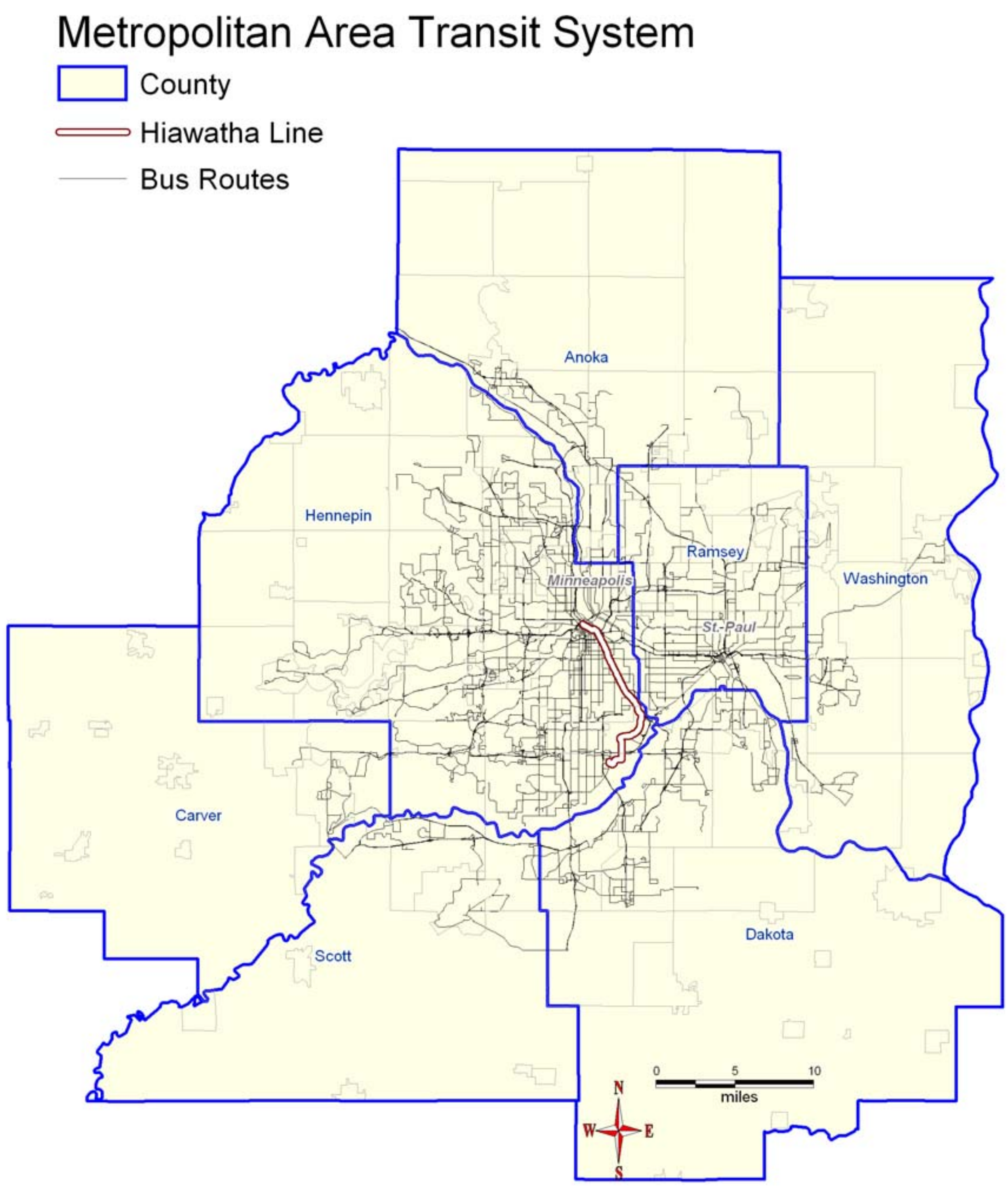
1. How does the cost of housing + transportation differ given Local Environment and Household characteristics - and what are those costs?
2. What locations can be considered affordable when the housing and transportation costs are considered?

One of the advantages to building the model using Local Environment and Household characteristics independent of each other is that it allows for increased functionality. First, it allows us to see the housing and transportation costs of an area given its population's current median income and average household size. Second, because the model treats these two types variables independently, we can input different household characteristics to determine the costs for a given household for all census block groups. This is particularly important in identifying areas that are affordable to low and moderate income working families in the Minneapolis-St. Paul MSA.

Minneapolis St. Paul Metro Area Transit System

The following map shows the transit system in the metropolitan area and can be used as a reference in understanding the importance of this system to the cost of transportation and housing. We have left this network off the other maps in this section because it visually obscures the underlying themes.

Table 27: Map of Transit System



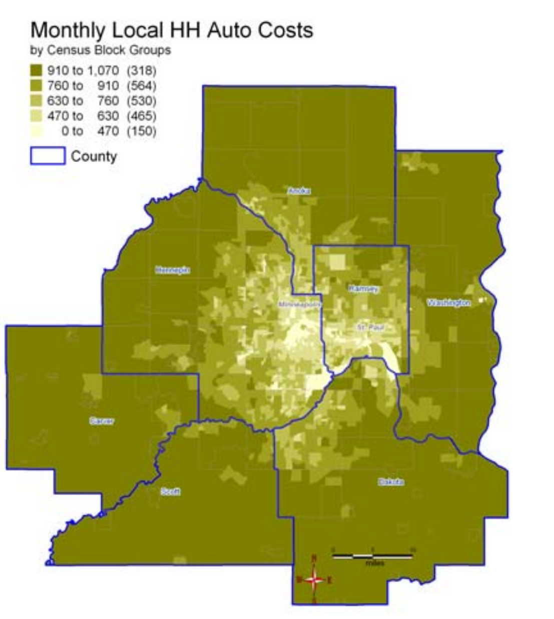
Modeled Transportation Cost Components

These maps depict the combined housing and transportation cost affordability for two household types: the households local to that Census block group (Local HH), and a moderate income household (Moderate Income HH):

- The Local HH maps are based on the current income and household size of the actual households that reside within a particular census block group, as of the 2000 U.S. Census. These maps depict how affordable the current location is for the existing households in that location.
- The Moderate Income HH maps are based on a household earning 80% of the Minneapolis-St. Paul MSA median income, which is $43,443, and a household size of 3 (the average household size in the MSA is 2.53). The Moderate Income HH maps identify what neighborhoods are affordable to this prototypical moderate-income household.

Using these two general categories, we map the Housing and Transportation costs separately and combined.

Table 28: Maps of the Modeled Transportation Costs Components



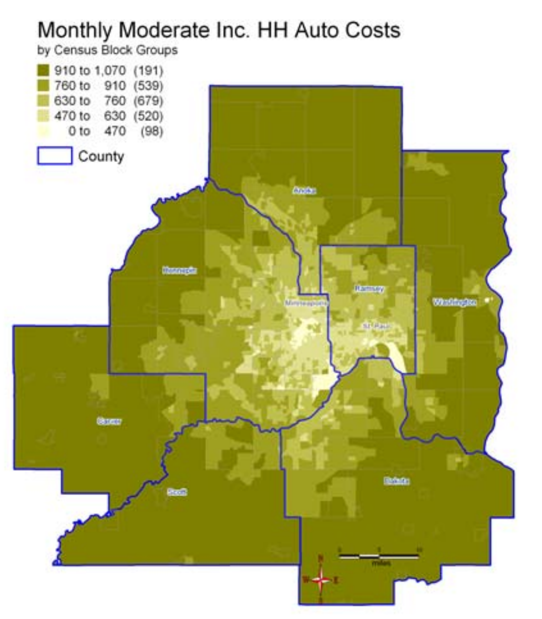
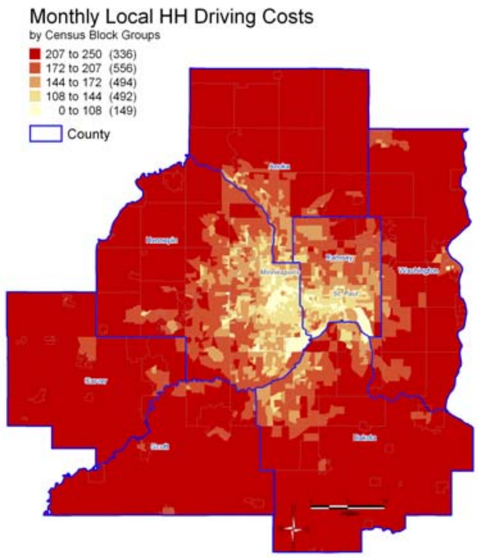
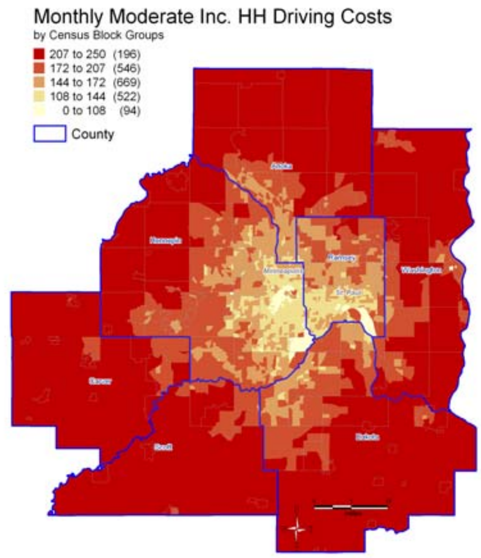
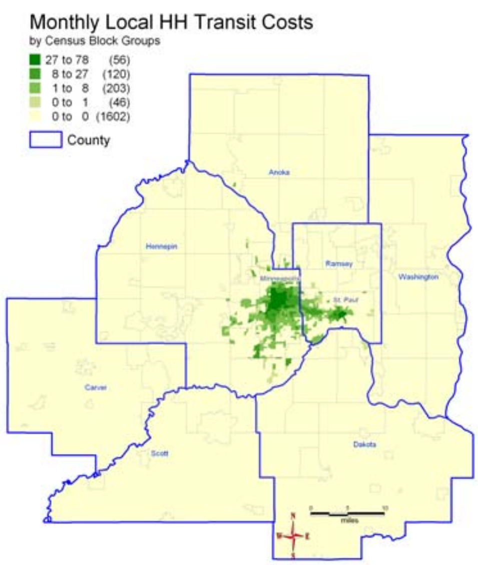
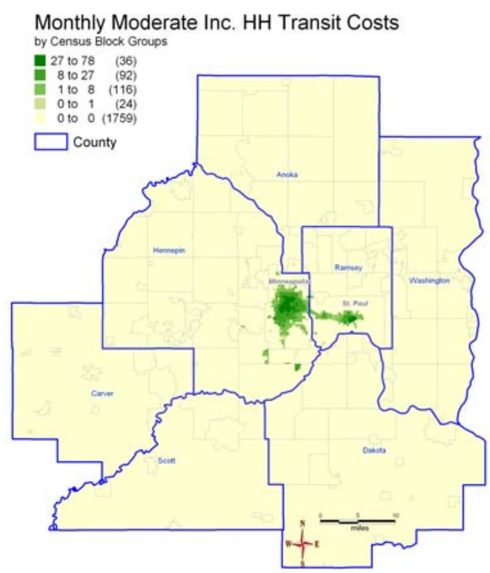


Table 28: Maps of the Modeled Transportation Costs Components





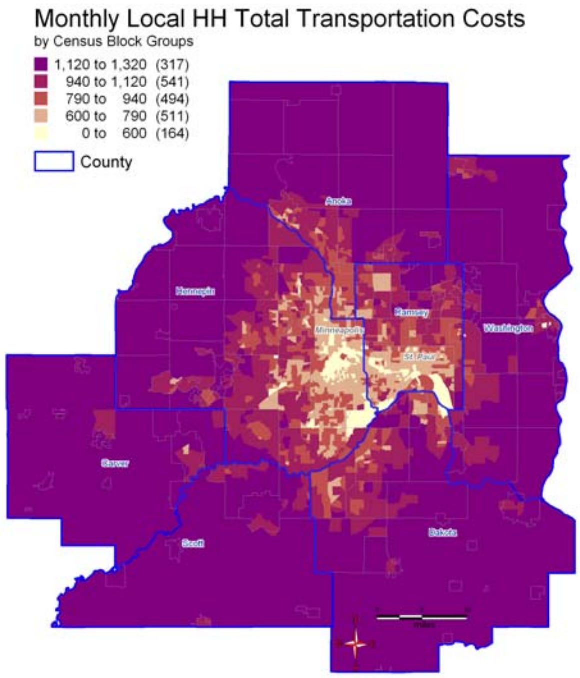


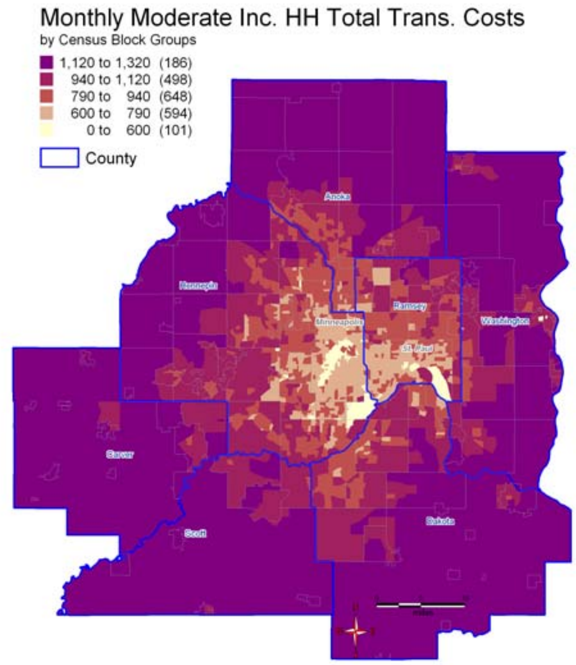


Modeled Total Transportation Costs

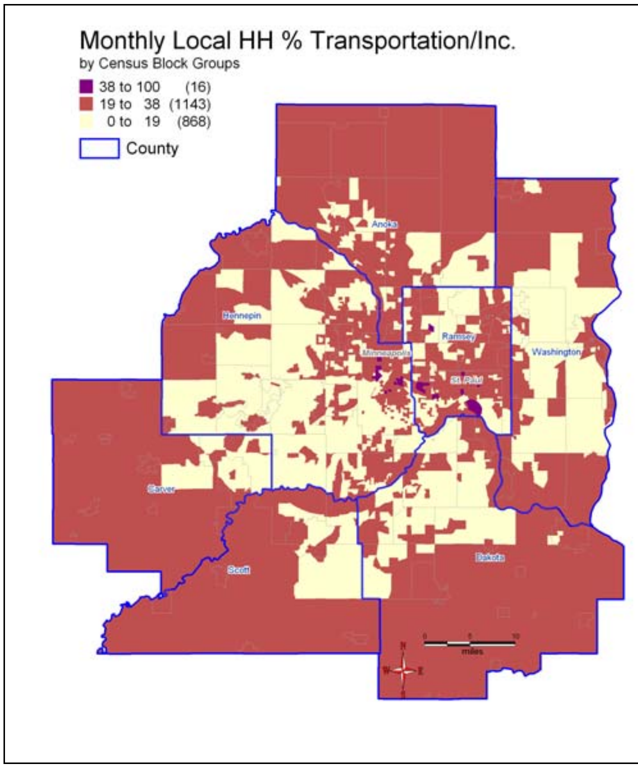
The next four maps show the total monthly cost of transportation (the sum of the three components as modeled above) for each of the two household groups; the Local Households and the moderate-income households, as a total expenditure in real dollars and as a percentage of income.[16]

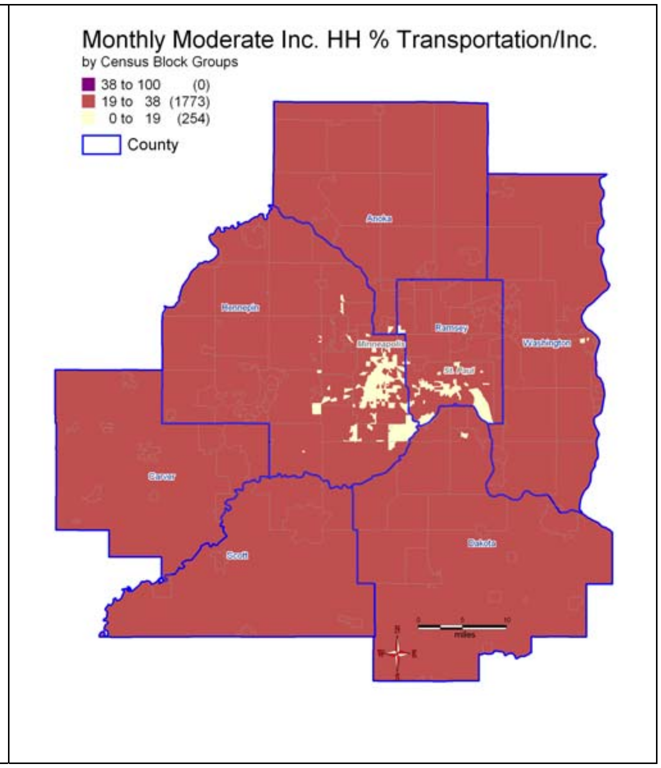
Table 29 Modeled Total Transportation Costs





[16] Note: the national average for percentage of household expenditures on transportation is 19.1% as of the Bureau of Labor Statistics 2003 Consumer Expenditure Survey.

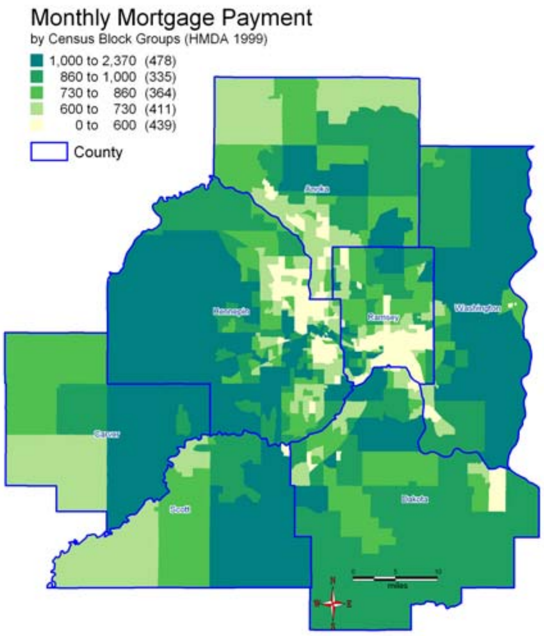


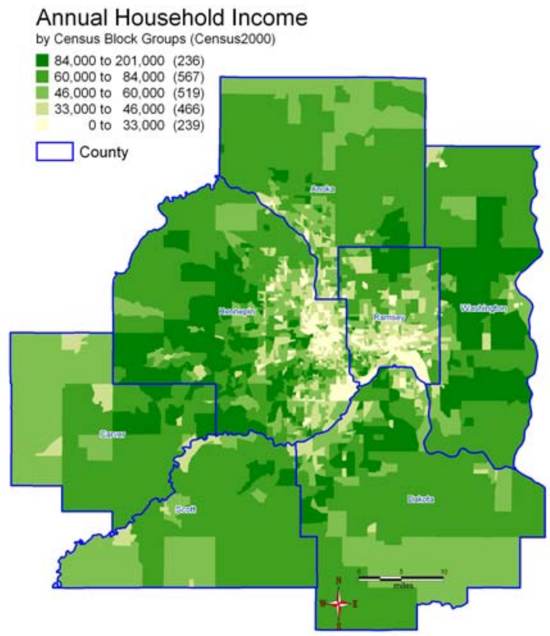


Modeled Housing Costs

In the housing series of maps, the first two maps show the Annual Household Income and average Monthly Mortgage Payment based on 1999 HMDA data. These two maps show that incomes and housing costs are generally parallel; areas with a higher income also have a tendency to have more expensive housing, and therefore higher mortgage payments.

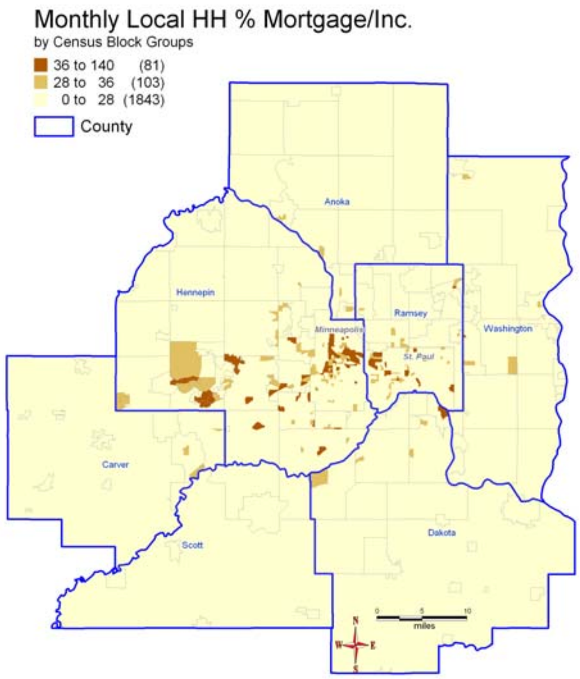
Table 30: Average Monthly Mortgage Payment in Comparison to Average Household Income

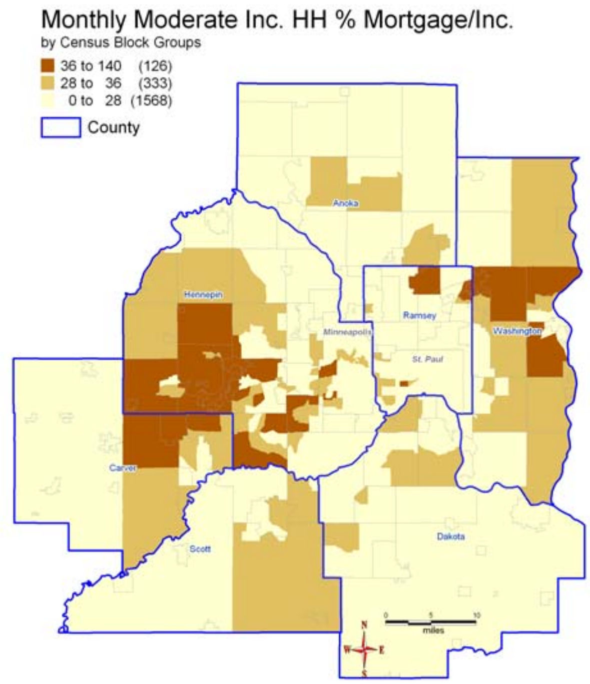




The following two maps show the actual monthly mortgage payment given the actual sales value of homes in the area and how affordable it is to the Local Households and the Moderate Income Households.

Table 31: Monthly Mortgage Payment Mapped as a Percentage of Income



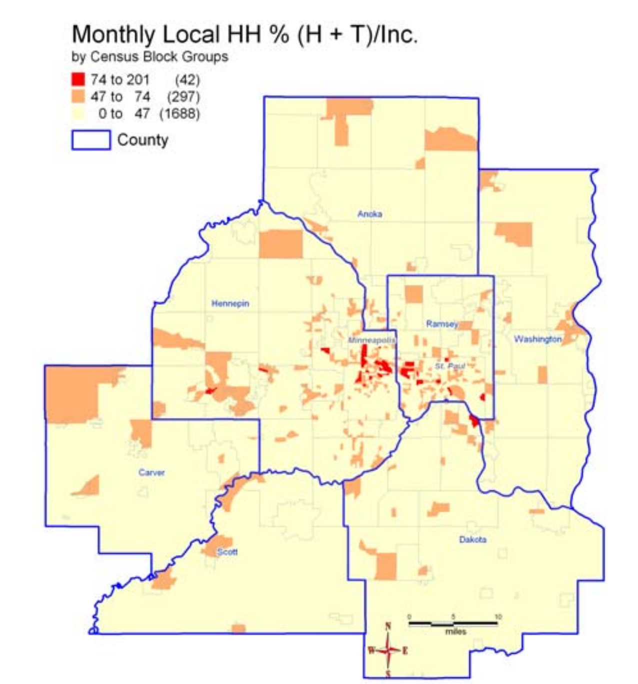


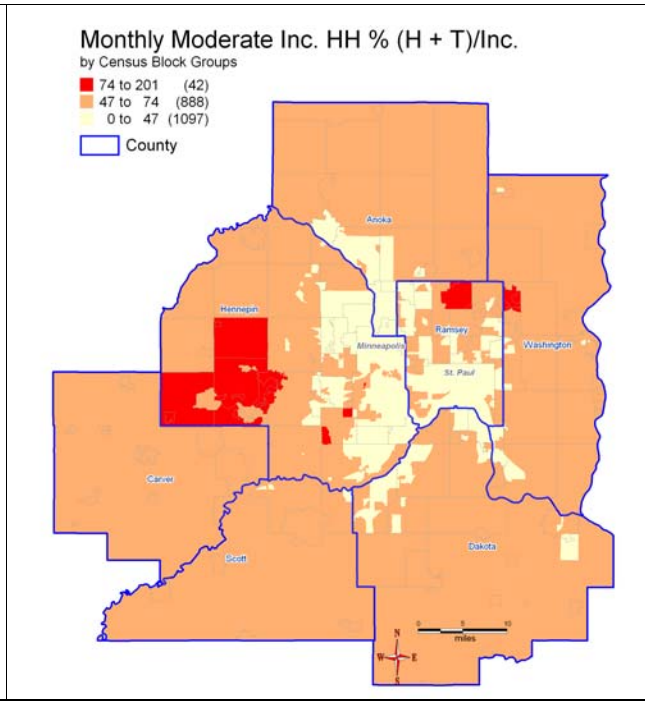
Modeled Housing + Transportation Costs- The Affordability Index

These final maps depict the Affordability Index! Using the same scale, both maps show the percent of income spent on the combined costs of transportation and owner-occupied housing in the Minneapolis-St. Paul region for the two household groups; current households in the block group and households earning 80% of the Area Median Income. To simplify the analysis at this larger regional level, the maps break affordability into three categories: 0-47% is affordable; 47-74% is moderately affordable to unaffordable, and 74% or more is extremely unaffordable.

In general, for households currently living in each block group, the majority of them live in a place that is affordable to them, given their income. This is indicated by the predominance of the light yellow color on the map. However, there are still a considerable number of census block groups that are in the moderately affordable to unaffordable class for "H+T" although they are affordable when just housing is considered. (See the maps in Table 28 for this comparison). Additionally, there are also areas in central Minneapolis that have a high degree of transit access, but are in the extremely unaffordable category for the combined costs of housing and transportation. This is likely an indication that the housing prices have escalated at a rate more quickly than incomes in those block groups since these same block groups are no longer considered unaffordable in the moderate income household map which depicts affordability for households earning 80% of the AMI.

Table 32: Maps of the Affordability Index (Combined Costs of Housing and Transportation as a Percentage of Income) for the Local Households and Moderate Income Households





These results indicate a substantial amount of variation in affordability, both in terms of the transportation costs and housing costs, and to whom they are affordable – those households currently living in the area, or moderate income households. The affordability does not follow a standard pattern and may even be of surprise to those familiar with the area. This is the benefit of this model-- the ability to show subtle to major cost differences associated with each neighborhood throughout the region, and to adapt it to various income levels.

The moderate income maps especially show the importance of looking at the total affordability of a place, beyond just housing costs. When transportation costs are considered along with housing costs, large areas outside of Minneapolis and St. Paul are unaffordable to families earning $43,443, or 80% of the median income in the MSA, which was $54,304 in the 2000 Census. But within these larger areas, it is just as important to consider the differences among the neighborhoods. Take for example, Dakota County. If the numbers were reported on a countywide level it would be unaffordable when transportation costs were factored. Yet, there are a number of census block groups that are affordable in this county. These affordable areas may not otherwise have been apparent to a home buyer without a model that accounts for the variation of transportation costs at a neighborhood scale.

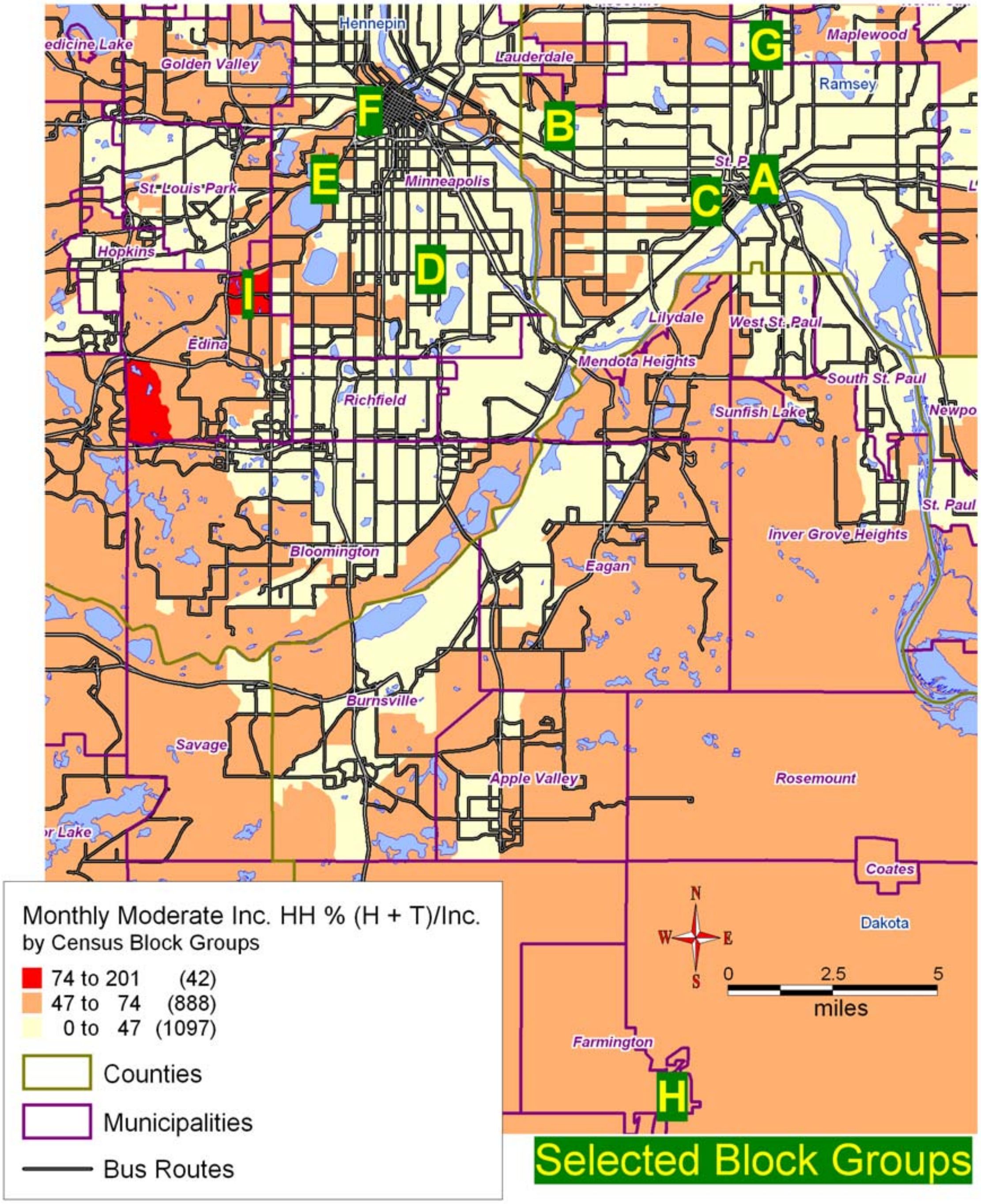
Sample Results

The Affordability Index is best illustrated by comparing the components in the index for several small geographies. The following tables show the components of the Affordability Index for 9 different locations in the Minneapolis-St. Paul Region. To show a range of Housing, Transportation, and Housing + Transportation costs, we selected three block groups in each of the two major cities; St. Paul and Minneapolis, and a block group in each of three suburban areas; Maplewood, Farmington, and Edina. The tables show these costs for the households currently living in the selected block groups, “Local HH”, and households earning 80% of the Area Median Income, “Moderate Inc. HH”. Additionally, we selected the block groups to represent a range of incomes. In Minneapolis and St. Paul, we selected a high, moderate, and low income block group, and the three suburbs each represent a low, moderate, or high income place. Table 31 displays the median income levels and average housing costs in these 9 block groups. A map with each of these areas identified by a letter ID corresponding to the ID in the table, shows these areas in relation to municipal boundaries the region’s bus route system. The region’s new light rail Hiawatha Line is not included on the map since the data pre-date the system.

Table 33: Income and Mortgage Payments for Selected Block Group Results

ID	*Block Group*	*City*	*Median Household Income*	*Average 1999 Mortgage Payment*
A	271230346013	St. Paul	$40,189	4516
B	271230319001	St. Paul	$15,347	$947
C	271230350001	St. Paul	$35,718	4829
D	270530110002	Minneapolis	$51,071	$662
E	270531066003	Minneapolis	$125,560	$1,975
F	270531052003	Minneapolis	$31,159	$970
G	271230422011	Maplewood	$43,313	$630
H	270370609031	Farmington	$61,127	$811
I	270530236002	Edina	$119,167	$2,142

Figure 18. Location of Nine Selected Block Groups to illustrate Affordability Index Results



Note the selected places are distributed from the urban core, to the inner-suburbs (G and H), to one outer suburb (H).

The next table presents the three components of the modeled transportation costs, plus total transportation costs, for the two proxy household types in each of these 9 block groups in order to represent a different type of local environment with distinct transportation costs.

Table 34: Modeled Transportation Costs for Households in 9 Different Block Groups

Transportation Costs for Current Households in the Selected Block Groups					
ID	*City Block Group*	*Modeled Auto Cost*	*Modeled VMT Cost*	*Modeled Transit Cost*	*Total Transportation Cost*
A	St. Paul	$232	$56	$60	$347
B	St. Paul	$417	$97	$8	$522
C	St. Paul	$398	$90	$24	$512
D	Minneapolis	$605	$137	$-	$742
E	Minneapolis	$666	$154	$6	$826
F	Minneapolis	$169	$41	$77	$288
G	Maplewood	$741	$169	$-	$910
H	Farmington	$994	$227	$-	$1,220
I	Edina	$798	$186	$-	$985
Transportation Costs for Households earning 80% of the Area Median Income					
ID	*City Block Group*	*Modeled Auto Cost*	*Modeled VMT Cost*	*Modeled Transit Cost*	*Total Transportation Cost*
A	St. Paul	$241	$60	$48	*$349*
B	St. Paul	$597	$141	$-	*$738*
C	St. Paul	$424	$99	$13	*$535*
D	Minneapolis	$592	$135	$-	*$727*
E	Minneapolis	$591	$136	$3	*$729*
F	Minneapolis	$187	$47	$65	*$299*
G	Maplewood	$744	$171	$-	*$915*
H	Farmington	$941	$214	$-	*$1,155*
I	*Edina*	*$706*	*$163*	*$-*	*$869*

The above transportation costs reflect the urban amenities and conveniences, or lack thereof, of a block group's location. The highest transportation costs are in Farmington (H), which is several miles from the urban core. In contrast, site F in Minneapolis exemplifies a transit rich place resulting in modeled auto costs of only $169 for Local Households and $187 for Moderate

Income Households. This area has a very rich network of buses and other amenities, including access to high job density and small block sizes.

In the next table, the separate and combined costs of housing and transportation are displayed as a percentage of income for the two household groups.

Table 35: Housing, Transportation, and Combined Housing and Transportation Costs as a Percentage of Income for Households in 9 Different Block Groups

		Local Households				*Moderate Income Households (80% AMI)*		
ID	**City Block Group**	**Median HH Income**	**T as % of Income**	**H as % of Income**	**(H+T) as % of Income**	**T as % of Income**	**H as % of Income**	**(H+T) as % of Income**
A	St. Paul	$40,189	10%	15%	26%	10%	14%	24%
B	St. Paul	$15,347	41%	74%	115%	20%	26%	47%
C	St. Paul	$35,718	17%	28%	45%	15%	23%	38%
D	Minneapolis	$51,071	17%	16%	33%	20%	18%	38%
E	Minneapolis	$125,560	8%	19%	27%	20%	55%	75%
F	Minneapolis	$31,159	11%	37%	48%	8%	27%	35%
G	Maplewood	$43,313	25%	17%	43%	25%	17%	43%
H	Farmington	$61,127	24%	16%	40%	32%	22%	54%
I	Edina	$119,167	10%	22%	31%	24%	59%	83%

When the housing and transportation costs are combined, Site F is still affordable to households earning 80% of the median income and, at 48% of income, is also moderately affordable to the current residents in the block group, who earn on average 57% of the MSA median income. Even though housing is 37% of their income, the lower transportation costs, 11% of income, maintain the affordability of this place.

On the other end of the spectrum is Site H, where transportation costs are over $1,000 dollars per month, and 24% and 32% of income for the local and moderate income households. The impact of these transportation costs is apparent when H + T costs are considered, which total 40% for the Local Household and 54% for a Moderate Income household. Housing costs are relatively lower in this area, but the high transportation costs negate the savings on the housing. In fact, the households currently living in Farmington spend more of their income on transportation than they do on housing.

Site D is also interesting, since the housing is affordable to both income groups, but the transportation is less affordable. The combined rate is still affordable to both these income groups, which in this case are those earning nearly the median income, $51,071, and those earning 80% of the median income. However, additional transit service, and possibly more jobs or other amenities, could further lower this areas transportation costs. Lower transportation costs would leverage the low housing costs in this area making this area affordable to an even broader range of incomes.

Summary

The model results illustrate the importance of accounting for the cost of transportation when measuring the affordability of a place. Because housing is intrinsically tied to a location, the associated locational costs are not easily separated. However, they are variable and can be influenced by the presence of transit along with proximity to employment, services, and walkable blocks. The Affordability Index model, by including several variables that represent the characteristics of the place, as well as household characteristics, and by breaking transportation costs into three components; auto ownership, auto use, and transit use, is able to account for and reflect this variation.