Location Affordability Index
Data and Methodology – Version 2.1

(September 2016)

Contents

Figures

Tables

Introduction

There is more to housing affordability than the size of your monthly rent or mortgage payment. Transportation costs are the second-biggest budget item for most families, but until recently there hasn't been an easy way for people to fully factor transportation costs into decisions about where to live and work. The goal of the Location Affordability Portal (LAP), launched in November 2013 by the U.S. Department of Housing and Urban Development (HUD) and the U.S. Department of Transportation (DOT), is to provide the public with reliable, user-friendly data and resources on combined housing and transportation costs to help consumers, policymakers, and developers make more informed decisions about where to live, work, and invest.

The LAP connects users to the Location Affordability Index, a robust, standardized data set containing household housing and transportation cost estimates at the Census block-group level for all 50 states and the District of Columbia. These estimates are generated using the Location Affordability Index Model (LAIM) Version 2.1, a combination of statistical modeling and data analysis primarily using data from a number of federal sources. The following documentation describes the methodology of current site update (LAIM Version 2.1) that went into effect in September 2016, highlighting the small but important improvements that have been put into place. A comprehensive account of the development of LAIM Version 2 can be found in Appendix A.

HUD and DOT are committed to engaging with the public to continually improve and expand this resource. Please email locationaffordability@hud.gov with any questions or comments.

Version History

LAIM Version 1 estimated three variables for transportation behavior (auto ownership, auto use, and transit use) as well as housing costs for homeowners and renters using separate Ordinary Least Squares (OLS) regression models. OLS assumes that structural errors are uncorrelated to each other. One weakness of this approach, though, was that the endogenous variables were correlated with the error term, making the modeled estimates of auto ownership, housing costs, and transit usage inconsistent and biased.

LAIM Version 2 represented a significant a methodological and technical advance from LAIM Version 1 in addition to updating all input data. Simultaneous (or structural) Equation Models (SEM) are the optimal statistical method to use when endogenous variables are all correlated with the error term. In SEM model, the endogenous variables are expressed as the function of predictor variables, other endogenous variables, and error.

LAIM Version 2 modelled auto ownership, housing costs, and transit usage for both homeowners and renters concurrently using SEM to account for the interrelationship of these factors.[1] The inputs to the SEM model include these six endogenous variables and 18 exogenous variables. As with Version 2, the

[1] Limitations of the data for VMT did not allow for its inclusion in the SEM; it continues to be modeled in Version 2 using OLS.

SEM model is used to estimate housing and transportation costs for eight different household profiles, in order to focus on the impact of the built environment on these costs by holding demographic characteristics constant.

LAIM Version 2.1 uses the same modeling methodology and data sources as Version 2 but uses the updated vintage and significantly improves the construction of the gravity models used to generate several variables. This data extraction process with detailed documentation will be helpful for the researchers to easily include or exclude variables, as may be required with various modifications to the statistical models.

LAIM Data Sources and Variables

I. Geographic Unit of Analysis and Scope

LAIM Version 2.1 is constructed at the Census block-group level using the 2010-14 American Community Survey (ACS) 5-year estimates as the primary dataset. Starting with LAIM Version 2, the Index covers the entire populated area of the 50 states and the District of Columbia.[2]

II. Basic Index Structure

LAIM Version 2.1 employs an SEM regression analysis for auto ownership, transit use, and housing costs and a second-order flexible form of ordinary least squares (OLS) model for VMT. It allows for all of the input variables to be used in the calculation of the coefficients. This somewhat complex modeling technique is employed to better model interactions between the endogenous variables. The goodness of fit is now measured by a combination of measures rather than by a simple R-squared value (see Section V. "Model Structure and Formula," IB. on goodness of fit measures on page 21 for further discussion – Kenny, D.A).

III. Data Sources

LAIM Version 2.1 is produced from data drawn from a combination of the following sources:

- U.S. Census American Community Survey (ACS) – an ongoing survey that generates data on community demographics, income, employment, transportation use, and housing characteristics. 2010-2014 survey data are used in LAIM Version 2.1.
- U.S. Census TIGER/Line Files – contains data on features of the natural and built environment such as roads, railroads, and rivers, as well as legal and statistical geographic areas.
- U.S. Census Longitudinal Employment-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) – detailed spatial distributions of workers' employment and residential locations and the relation between the two at the Census Block level, including characteristic detail on age, earnings, industry distributions, and local workforce indicators (see overview). LODES and OnTheMap Version 7.2, which are built on 2010 Census data, are used here.
- U.S. Census ZIP Code Tabulation Areas (ZCTAs) – are generalized areal representation of United States Postal Service (USPS) ZIP Code service areas for 2010.
- U.S. Regular Gasoline Prices - "Petroleum & Other Liquids" data (U.S. Department of Energy, Energy Information Administration) contains data on weekly retail gasoline and diesel prices, by

[2] Excluding the few uninhabited block groups in the United States for which demographic data is not available. In addition, there is unfortunately insufficient data to create the Index for U.S. territories.

Petroleum Administration for Defense Districts (PADD) district and sub district. 2012 gas prices were used in LAIM Version 2.1.

- National Household Travel Survey (NHTS) (2009) – survey sponsored by FHWA collects data on both long-distance and local travel by the American Public. National driving records were obtained for the entire country from NHTS.
- Odometer Data for Illinois Urbanized Areas – the Illinois Environmental Protection Agency (IL EPA) provided Illinois odometer data for 2010 and 2012 with details on VIN, test date and time, odometer reading, zip code, model, and year.
- Census Home-To-Work Tract Flows– part of the Census Transportation Planning Products 2006-10 tabulations, which are built on ACS data from the same period. Home-To-Work flow data was used to estimate the 2012 employment for every block group of Massachusetts.
- Massachusetts Employment and Wages (ES-202) – county-level employment data for Massachusetts was obtained for 2012.
- National Transit Data (NTD) – data was obtained for 2012 to calculate transit costs. Contains data on transportation revenue by transit agency, urban area, and mode and type of service.
- Census 2010 Census Tract Relationship Files – files with population and housing unit information and area measurements.

LAIM Version 2.1 represents a significant improvement in using the updated data sources. The data sources with the updated vintages used in LAIM Version 2.1 is shown in Table 1: Model Data Sources with Vintage.

Table 1: Model Data Sources with Vintage

Name	LAIM Version 2	LAIM Version 2.1
American Community Survey (ACS)	2008-2012	2010-2014
LEHD Origin-Destination Employment Statistics (LODES)	OnTheMap Version 7	OnTheMap Version 7.2
Gasoline Prices	2008	2012
National Transit Database	2010	2012
National Household Travel Survey	2009	2009
Massachusetts ES202	2010	2012
Vehicle Miles Traveled (VMT)	2008-2010	2010-2012

These data describe relevant characteristics of every census block group in the United States. Census block groups contain between 600 and 3,000 people and vary in size depending on an area's population density. They range from only a few city blocks to significant proportions of some rural counties. Block groups are the smallest geographical unit for which reliable data is available; they can generally be thought of as representing neighborhoods.

IV. Data Extraction

The data extraction process includes several script files written in the F# computer language, each of which will yield one or more data files formatted as comma-separated value (CSV) files. These resultant data files are subsequently loaded in relational database and then joined together on common primary keys and further formatted using a final script file written in a dialect of Structured Query Language (SQL), with the final result then being able to be output in any of a number of formats. The details of data extract process narrative are included as a separate document.

The data extraction procedure utilized in the current method has several advantages over the previous methodology.[3]

1. **ACS Data Extraction**: All information necessary to build all of the ACS-related intermediate data files (e.g., block group, tract, county or Combined Base Statistical Area) is included in the single script. The generalized design of the new data structure allows for any new summation levels necessary for the purposes of the model (e.g., state, CSA, NECTA etc.). The existing ACS summation level scraper workflows can be easily modified to include additional sequence variables or to exclude them as required by modifications to the model.
2. **Gravity Extraction**: Employment gravities are calculated at the block level and summed to the block group level, as opposed to the previous process of only calculating them at the block-group level. This improvement allows the gravity model to capture inter-block and intra-block-group employment flows, lending increased weight to such local employment patterns.
3. **Distance calculation**: The distances between the origin and destination blocks are calculated using a haversine spherical approximation. This method improves on the Euclidean distance approximation employed in LAIM Version 2 and has a far greater margin of error when calculating distances over the earth, reducing the overall level of noise in the local employment data.

IV. Variables

To design the SEM for LAIM Version 2, the team started with a pool of potential independent variables (referred to in the SEM as *exogenous variables*) representing all of the possible influences on housing and transportation costs for which data were available. They then selected variables for use in the model according to the strength of their correlation with the dependent (endogenous) variables and their statistical significance, drawing on the economic theoretical framework developed with federal

[3] For complete documentation of modeling code and data extraction process narrative, please see http://www.locationaffordability.info/downloads/ModelingCode.pdf

stakeholders and the site's Technical Review Panel (which met five times between November 2011 and June 2013). Table 2: Overview of LAIM Version 2.1 Variables lists the set of variables used in LAIM Version 2.1, with endogenous variables shaded (see II.C. in Appendix A "Variable Selection" for how the final set variables was chosen). The right-hand side of the SEM equation contains both exogenous and endogenous variables with six equations and the determinants of all shaded endogenous variables are explicitly modeled.

Table 2: Overview of LAIM Version 2.1 Variables

Input	Description	Data Source
Gross Density	# of households (HH) / total acres	Census ACS, TIGER/Line files
Block Density	# of blocks / total land area	Census TIGER/Line files
Employment Access Index	Number of jobs in area block groups / squared distance of block groups	Census LEHD-LODES
Retail Employment Access Index	Number of retail jobs in area block groups / squared distance of block groups	Census LEHD-LODES
Median Commute Distance	Calculated from data on spatial distributions of workers' employment and residential locations and the relation between the two at the Census block level	Census LEHD-LODES
Job Density	# of jobs / total land area	Census LEHD-LODES
Retail Density	# of retail jobs / total land area	Census LEHD-LODES
Fraction of Rental Units	Number of rental units as a percentage of total housing units	Census ACS
Fraction of Single Family Detached Housing Units	Number of single family detached housing units as a percentage of total housing units	Census ACS
Median Rooms/Owner HU	Median number of rooms in owner occupied housing units (HU)	Census ACS
Median Rooms/Renter HU	Median number of rooms in renter occupied housing units	Census ACS
Fraction of Median Income Owners	Median income for owners at the block group level as a percentage of either CBSA or County median income (County for rural areas / CBSA for Metropolitan and Micropolitan Areas)*	Census ACS
Fraction of Area Median Income	Median income for renters at the block group level as a percentage of either CBSA or County median income	Census ACS

Input	Description	Data Source
Renters	(County for rural areas / CBSA for Metropolitan and Micropolitan Areas)	
Average Household Size: Owners	Calculated from data on Tenure and Total Population in Occupied Housing Units by Tenure	Census ACS
Average Household Size: Renters	Calculated from data on Tenure and Total Population in Occupied Housing Units by Tenure	Census ACS
Average Commuters per Household Owners	Calculated using the total number of workers 16 years and over who do not work at home	Census ACS
Average Commuters per Household Renters	Calculated using the total number of workers 16 years and over who do not work at home	Census ACS
Median Selected Monthly Owner Costs	Includes mortgage payments, utilities, fuel, and condominium and mobile home fees where appropriate	Census ACS
Median Gross Rent	Includes contract rent as well as utilities and fuel if paid by the renter	Census ACS
Autos per Household Owners	Calculated from Aggregate Number of Vehicles Available by Tenure and Occupied Housing Units	Census ACS
Autos per Household Renters	Calculated from Aggregate Number of Vehicles Available by Tenure and Occupied Housing Units	Census ACS
Percent Transit Journey to Work Owners	Calculated from Means of Transportation to Work by Tenure	Census ACS
Percent Transit Journey to Work Renters	Calculated from Means of Transportation to Work by Tenure	Census ACS

*CBSA = Core Based Statistical Area

The following detailed descriptions of variables used for LAIM Version 2.1 are organized according to the seven most significant factors that influence transportation costs: population density; street connectivity and walkability; employment access and diversity; housing characteristics; resident household characteristics; housing costs; and household travel behavior.

A. Population Density

1. Gross Household Density

Population density has been found to be one of the largest factors in explaining the variation in all three transportation dependent variables (Ewing et al. 2003). Various definitions of density have been constructed and tested, with Gross Household Density emerging as the favored measure for modeling both housing and transportation costs. Gross Household Density is calculated as total households (from the ACS) divided by total land acres (calculated using TIGER/Line files).

B. Street Connectivity and Walkability

2. Block Density

Measures of street connectivity have been found to be good proxies for pedestrian friendliness and walkability. Greater connectivity created by numerous streets and intersections creates smaller blocks and tends to lead to less dependence on automobiles as well as shorter average auto trips, and more use of transit. While other factors (e.g. public safety) clearly have an impact on the pedestrian environment, these types of street connectivity measures have been found to be an important driver of auto ownership, auto use, and transit use. The LAIM uses average Block Density, which is defined as the total block group land area (in acres) divided by the number of blocks within a block group.

C. Employment Access and Diversity

Employment numbers are calculated using *OnTheMap* Version 7.2 which provides Longitudinal Employer-Household Dynamics (LEHD) Origin Destination Employment Statistics (LODES) at the Census block group level.

Unfortunately, these data are currently unavailable in Massachusetts, requiring the creation of a synthetic data set by processing and merging together data from several other preexisting data products. Table 3 lists the input data sets and resulting output data from processing.

Table 3: Data Inputs and Post-Processing Outputs

Input Data set	Output data
Tract-level "A20210—Industry (Workers 16 years and over)" file for Massachusetts from 2006–10 CTPP	tract IDs, county IDs, and the fractions of total county jobs and county retail jobs in each constituent tract
Massachusetts' 2012 ES-202 county employment data	county IDs and the total and retail average monthly employment of the county
2006–10 CTPP Census Tract Flows	residence and workplace tract IDs for those records whose residences or workplaces are in Massachusetts; the fraction of worker flow to the workplace tract within that tract's encompassing county.
2010 block-level Census Population & Housing Unit Counts (TIGER/Line Shapefiles with Selected Demographic and Economic Data)	block IDs, tract IDs, and the fraction of block's 2010 Census population within its encompassing block.
2012 block-level Census Population & Housing Unit Counts (TIGER/Line Shapefiles with Selected	block IDs, block group IDs, and the longitude and latitude coordinates of the block's internal

Demographic and Economic Data)	point.

These outputs were then joined together to yield a final, synthetic dataset containing the residence block group ID, block ID, and block's latitude and longitude coordinates; the workplace block's latitude and longitude coordinates; the estimated total job flow between the residence and workplace blocks; and the estimated numbers of total and retail jobs in the workplace block.

Measures of employment access and density provide not only an examination of access to work, but are good proxies for economic activity. While they overlap in what they measure, each have a unique aspect that make them more predictive when used in concert than when used individually: the average median commute distance captures only at the access to work, while the Employment Access Index looks at where jobs exist and captures economic activity and the local job density captures the count of actual jobs.

The gravity model employed in LAIM Version 2.1 represents an improvement over LAIM version 2 in two primary ways. First, the LAIM Version 2.1 gravity model concretely incorporates same-block group employment where the previous model was silent in regards to their treatment. In particular, the Version 2.1 gravity model assigns a nominal distance of one (1) mile to those jobs whose corresponding employees are located within the same block group, so as to incorporate their importance into the measures of local employment while also weighting them proportionately to the contributions of all other jobs. This is in contrast to both the Version 2 gravity model, which is silent as to the treatment of these same-block group jobs, as well as to the default Version 2.1 procedure, which is to measure distance from the population-weighted centroids of the block groups in question—which would yield a distance of zero (0) miles and would cause undefined behavior in the gravity model. Second, the LAIM Version 2.1 gravity model adjusts the methodology used for distance calculations from employees' homes to their corresponding employment locations from Euclidean distance (a planar method) to haversine distance (a spherical approximation) to minimize geodetic error. In particular, the haversine distance methodology was selected due to it being sufficiently precise over the maximum distances encountered for its low computational complexity—an important concern given the iterative nature of distance calculations in the gravity model.

3. Employment Access Index

The Employment Access Index is generated using a gravity model, which employs an inverse square formula to account for both the quantity of and distance to all employment destinations relative to any given block. It is calculated by summing the total number of jobs divided by the square of the total distance to those jobs. Relative to a simple employment density measure, the Employment Access Index gives a better measure of job opportunity, and thus a better understanding of job access, because the gravity model enables consideration of jobs both directly in and adjacent to a given block. This index also serves as a proxy for access to economic activity.

The Employment Access Index is calculated as:

$$E_b = \sum_{i=1}^{n} \frac{p_{E,i}}{r_i^2}$$

Where

E_b = employment gravity at the block level

$p_{E,i}$ = # of total jobs in the i^{th} block

r_i^2 = squared distance in miles between given block and i^{th} block

n = total number of Census blocks

Proportion of total workers in a block to total workers in the corresponding block group:

$$P_E = \frac{W_{E,b}}{W_{E,B}}$$

Where

P_E = proportion of total workers in a block to total workers in the corresponding block group

$W_{E,b}$ = # of total workers in a given block

$W_{E,B}$ = # of total workers in the block group that contains the given block

Employment Gravity at the block group level:

$$E_B = \sum_{i=1}^{n} P_{E,i} E_{b,i}$$

Where

E_B = employment gravity at the block group level

$E_{b,i}$ = employment gravity for the i^{th} block

$P_{E,i}$ = proportion of total workers in the i^{th} block to total workers in the given block group

n = total number of Census blocks within the given block group

As jobs get farther away from a given Census block group their contribution to the Employment Access Index is reduced; for example, one job a mile away adds one, but a job 10 miles away adds 0.01. All jobs in all U.S. Census block groups are included in this measure.

4. Retail Employment Access Index

This index is calculated using the same method as the Employment Access Index (above) only using the number of jobs in NAICS sector 44-45 (Retail Trade).

5. Median Commute Distance

Median commute distance is calculated using LODES data. Commute distances are calculated for each Census block using a haversine spherical distance approximation between the origin and destination Census blocks. Commute distance values for each block are then ordered by distance to obtain the median value for the block group of interest.

Using a haversine spherical distance approximation yields an improvement over a Euclidean distance employed in LAIM Version 2 as a Euclidean (planar) distance approximation has a far greater margin of error when calculating distances over the earth than a haversine spherical approximation. This improvement should work to reduce the overall level of noise in the local employment data because the error of planar approximation is not uniform and margin of error increases with increase in latitude.

6. Local Job Density

A block group's Local Job Density is its number of jobs per square mile. Using LODES data, the total number of jobs in the buffer is calculated and divided by the land area. The generic formula for this calculation for block group *i* is:

$$Local\ Job\ Density_i = \frac{Jobs\ in\ relevant\ geography}{Area\ of\ relevant\ geography\ (sq.mi.)}$$

Two different calculation procedures are used to determine local job density, both of which use as a reference area a half-mile buffer (Zhao, Chow, Li, Ubaka, & Gan, 2003) around the population-weighted centroid of each block group to account for the wide variability of block groups:

1. If the block group's land area is greater than the reference area, then the local job density equals the number of jobs in the block group divided by the block group's land area, in square miles (see Figure 1a below).
2. If the block group's land area is less than or equal to the reference area, then the local job density equals the total number of jobs divided by the total land area, in square miles, for both the block group in question as well as any other block groups whose population-weighted centroids are located inclusively within one-half mile of the population-weighted centroid of the block group in question (see Figure 1b).

Figure 1: Local Job Density Calculations

Block groups with data included in Local Job Density Calculation	Block groups not included in Local Job Density calculation	X Reference area with centroid for block group *i*	O Other block group centroids

Figure 1a

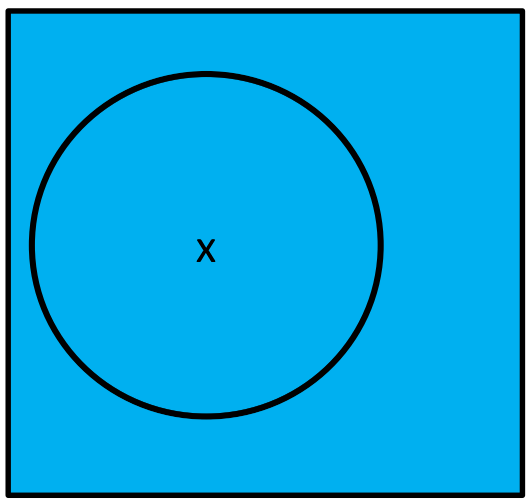
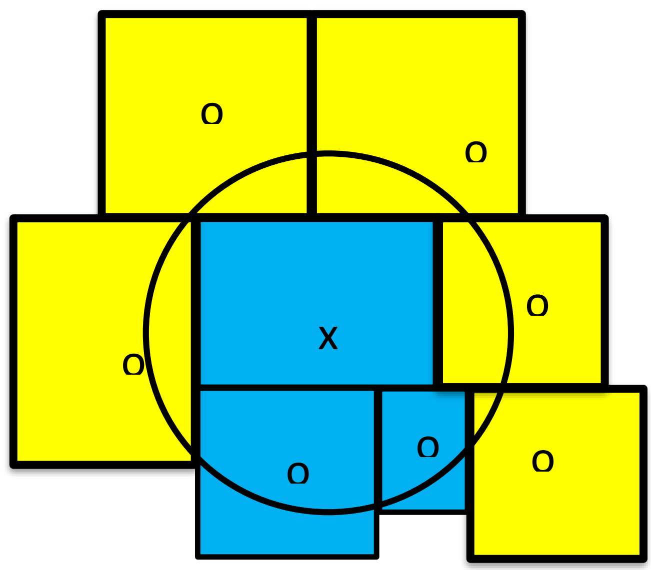


Figure 1b



7. Local Retail Job Density

Local Retail Job Density is calculated in the same way as Local Job Density, except using LODES data for retail jobs only.

D. Housing Characteristics

Certain characteristics of neighborhood housing stock and tenure have been shown to influence household travel behavior.

8. Fraction of Rental Units

The number of rental units as a percentage of total housing units is calculated using data on Tenure from the ACS.

9. Fraction of Single-Family Detached Housing Units

The number of single-family detached housing units as a percentage of total housing units is calculated using data on Tenure by Units in Structure from the ACS.

10. Median Number of Rooms in Owner-Occupied Housing Units

11. Median Number of Rooms in Renter-Occupied Housing Units

Median Number of Rooms for Owner- and Renter-Occupied Housing Units is calculated using 2010-2014 ACS data on Median Number of Rooms by Tenure and is included as an exogenous variable. In cases where block-group data for the Median Number of Rooms is suppressed, the value for the tract is used in running the model but excluded for the purpose of model calibration.

E. Resident Household Characteristics

The 2010-2014 ACS 5-year estimates serve as the primary data source for variables pertaining to resident household characteristics. For all of the following household characteristics variables, tract values are used in place of suppressed block-group-level data for running the model but excluded for the purpose of model calibration.

12. Area Median Income

Median household income is obtained directly from the ACS at the Core Based Statistical Area (CBSA) level for block groups in Metropolitan and Micropolitan area and at the county level for all block groups in counties not included in a CBSA (i.e. "noncore" counties).[4]

13. Fraction of Area Median Income Owners

14. Fraction of Area Median Income Renters

Fractions of area median income for owners and renters are calculated as the ratio of median income for owners or renters at the block group level to the Area Median Income (see paragraph E.12.).

15. Average Household Size Owners

16. Average Household Size Renters

Average Household Sizes for owners and renters are calculated by dividing the total population in owner or renter units by the number of owner or renter units, using Tenure and Total Population in Occupied Housing Units by Tenure to define the universes of Owner-Occupied and Renter-Occupied Housing Units (see paragraph E. iv).

17. Average Commuters per Household Owners

18. Average Commuters per Household Renters

Average commuters per household for owners and renters are calculated using the total number of workers 16 years and older who do not work at home from Means of Transportation to Work and Tenure to define Owner-Occupied and Renter-Occupied Housing Units. Because Means of Transportation to Work includes workers not living in occupied housing units (i.e., those living in group quarters), the ratio of Total Population in Owner-Occupied or Renter-Occupied Housing Units to Total Population is used to scale the count of commuters to better represent those living in households (see paragraph E. vi).

F. Housing Costs

[4] See here for an explanation of the difference between metropolitan, nonmetropolitan, and noncore counties: http://www.ers.usda.gov/topics/rural-economy-population/rural-classifications/what-is-rural.aspx.

The 2010-2014 ACS 5-year estimates serve as the data source for variables pertaining to housing costs.

19. Median Selected Monthly Owner Costs

Median Selected Monthly Owner Costs are taken directly from the ACS and include mortgage payments, utilities, fuel, and condominium and mobile home fees, where appropriate.

20. Median Gross Rent

Median Gross Rent is taken directly from the ACS and includes contract rent as well as utilities and fuel if paid by the renter.

G. Household Travel Behavior

The 2010-2014 ACS 5-year estimates provide source data for variables describing household travel behavior.

21. Autos per Household Owners

22. Autos per Household Renters

Autos per Household Owners and Autos per Household Renters are calculated from Aggregate Number of Vehicles Available by Tenure and Occupied Housing Units.

23. Percent Transit Journey to Work Owners

24. Percent Transit Journey to Work Renters

Although transit accessibility is a key correlate of transportation costs, transit service data is not ubiquitous and not universally available, even for areas with high-quality transit. As a result, the Index does not use direct measures of transit access as inputs. Instead, the model incorporates as proxies two endogenous variables—percent of commuters using transit for journey to work for home-owners and renters—which are robust indirect measures of transit use and have the benefit ubiquity across all urban, suburban, and rural settings.

Means of Transportation to Work by Tenure is used to calculate a percentages of commuters in owner-occupied and renter-occupied housing utilizing public transit.

V. Variable Transformations

Similar to LAIM Version 2, SEM variables are transformed to allow for better fits for non-linear relationships in LAIM Version 2.1. The approach is to apply a series of transformations to each of the endogenous and exogenous variables and pick the transformation that produces the most normal distribution for each one (i.e., the distribution that maximizes the R^2 value when compared with a normal distribution). This transformed variable is then standardized by subtracting the mean of the transformed distribution and dividing by the standard deviation:

$$x'_i = \frac{f(x_i) - \overline{f(x)}}{SD_{f(x)}}$$

Where

x'_i = the transformed and standardized value for a given observation of variable x

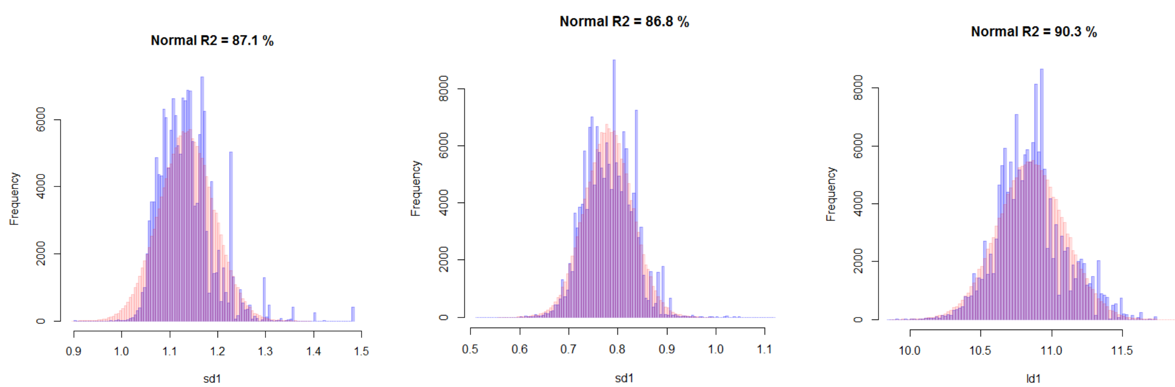
$f(x_i)$ = the transformed value for a given observation of variable x

$\overline{f(x)}$ = the mean of the transformed variable x

$SD_{f(x)}$ = the standard deviation of the transformed variable x

This standardization was applied for all the variables in the SEM function as listed in Table 4 (next page) to handle the wide variation in values. Please see Appendix II.B. for details on advantages of transformation of these variables. Figure 2 compares transformed distributions for several variables used in SEM model to normal distributions.

Figure 2: Distributions of SEM Variables over All Census Block Groups



Area Income Fraction Owners
Square Root - $\sqrt{x}$

Area Income Fraction Renters
Square Root - $\sqrt{x}$

Area Median Income
Natural Log - ln(x)

Table 4: Variables Used to Estimate the Model, with Transformations and Descriptive Statistics

Name	Transformation	Mean of Transformed Variables	Standard Deviation of Transformed Variables
1. Gross HH Density	Square Root	1.375	1.185
2. Block Density	Square Root	0.289	0.172
3. Employment Access Index	Natural Log	4.078	1.120
4. Retail Employment Access Index	Natural Log	1.777	1.311
5. Median Commute Distance	Natural Log	2.823	0.914
6. Local Job Density	Natural Log	-0.236	2.113
7. Local Retail Job Density	Square Root	0.406	0.531
8. Fraction Rental Units	Square Root	0.585	0.177
9. Fraction Single-Family Detached HU	Linear	62.570	27.465
10. Median Rooms/Owner HU	Linear	6.158	0.926
11. Median Rooms/Renter HU	Linear	4.662	1.025
12. Area Median Income	Natural Log	10.854	0.236
13. Area Income Fraction Owners	Square Root	1.135	0.057
14. Area Income Fraction Renters	Square Root	0.781	0.048
15. Average HH Size Owner	Natural Log	0.956	0.227
16. Average HH Size Renters	Natural Log	0.908	0.324
17. Average Commuters/HH Owners	Linear	1.227	0.153
18. Average Commuters/HH Renters	Linear	1.055	0.161
19. Median SMOC	Natural Log	7.217	0.379
20. Median Gross Rent	Natural Log	6.766	0.375
21. Autos/HH Owners	Linear	1.954	0.410
22. Autos/HH Renters	Linear	1.368	0.484
23. Percent Transit Journey to Work Owners	Linear	1.900	4.677
24. Percent Transit Journey to Work renters	Linear	5.827	13.559

J2W = Journey to Work HH = Households
HU = Housing Units SMOC = Selected Monthly Ownership Costs
Endogenous variables are shaded.

LAIM Structure and Formula

I. Simultaneous Equations Model

The SEM used in LAIM Version 2.1 consists of six nested equations, each drawing from a pool of 18 exogenous variables, that predict six interrelated endogenous variables. The standard form of SEM model and the distributional assumptions of error terms ar e provided in Appendix D.

A. SEM Structure

Table 5 (following page) shows the structure of the SEM model used in LAIM Version 2.1, organized by the six nested equations for the model's endogenous variables (the left-hand terms for which are shaded and bolded). All endogenous variables appearing as exogenous variables in other nested equations are shaded as well. The exogenous variables in each nested equation were selected based on the strength and statistical significance of their correlation with the endogenous variables. As discussed in variables section, exogenous variables were also selected based on household transportation behavior which is highly affected by household density, street connectivity and walkability, employment access and diversity, housing characteristics, and housing costs.

Table 5: Simultaneous Equations Model: Endogenous and Exogenous Variables

Variables	Estimate	Std. Error	Z-Value
Autos/HH Owners			
3. Employment Access	0.036	0.004	9.454
12. Area Median Income	0.039	0.003	13.408
17. Commuters/HH Owners	0.037	0.003	14.233
4. Retail Employment Access	-0.058	0.004	-16.463
8. Fraction Rental Units	0.073	0.003	27.959
13. Area Income Fraction Owners	-0.087	0.003	-34.532
19. Median SMOC	0.128	0.003	49.466
10. Median Rooms/Owner HU	0.118	0.002	58.213
1. Gross HH Density	-0.224	0.003	- 67.987
2. Block Density	-0.214	0.003	-75.303
9. Fraction Single Detached HU	0.220	0.003	82.717
15. HH Size Owner	0.329	0.002	171.574
Autos/HH Renters			
12. Area Median Income	0.019	0.003	6.442
4. Retail Employment Access	-0.028	0.004	-7.364
14. Area Income Fraction Renters	-0.024	0.003	-9.409
5. Median J2W Miles	0.039	0.002	17.063
3. Employment Access	0.074	0.004	17.999
2. Block Density	-0.094	0.003	-28.129
6. Local Job Density	-0.109	0.003	-34.630
18. Commuters/HH Renters	0.131	0.003	43.659
1. Gross HH Density	-0.178	0.003	-51.219
11. Median Rooms/Renter HU	0.150	0.002	60.786
9. Fraction Single Detached HU	0.179	0.002	72.108
20. Median Gross Rent	0.220	0.003	85.776
16. HH Size Renters	0.207	0.002	93.731
Median SMOC			

Variables	Estimate	Std. Error	Z-Value
5. Median J2W Miles	0.009	0.002	4.111
4. Retail Employment Access	-0.023	0.003	-6.582
6. Local Job Density	-0.042	0.003	-14.513
3. Employment Access	0.119	0.004	31.526
2. Block Density	-0.124	0.003	-40.375
15. HH Size Owner	0.082	0.002	44.389
1. Gross HH Density	0.162	0.003	49.951
17. Commuters/HH Owners	-0.137	0.003	-54.479
9. Fraction Single Detached HU	-0.162	0.003	-63.916
8. Fraction Rental Units	-0.189	0.003	-74.647
10. Median Rooms/Owner HU	0.214	0.002	112.706
13. Area Income Fraction Owners	0.338	0.002	144.097
12. Area Median Income	0.591	0.002	241.722
Median Gross Rent			
2. Block Density	-0.015	0.003	-5.538
3. Employment Access	-0.021	0.004	-5.796
4. Retail Employment Access	0.104	0.003	31.332
8. Fraction Rental Units	-0.073	0.002	-34.069
5. Median J2W Miles	-0.077	0.002	-39.199
1. Gross HH Density	0.116	0.003	39.627
16. HH Size Renters	0.139	0.002	72.403
14. Area Income Fraction Renters	0.158	0.002	86.221
11. Median Rooms/Renter HU	0.228	0.002	107.949
12. Area Median Income	0.257	0.002	122.833
19. SMOC	0.372	0.002	171.774
Transit %J2W Owners			
7. Local Retail Job Density	0.009	0.002	4.156
4. Retail Employment Access	-0.046	0.004	-11.966

Variables	Estimate	Std. Error	Z-Value
2. Block Density	0.042	0.003	13.700
3. Employment Access	-0.058	0.004	-14.530
24. Transit %J2W renters	0.161	0.007	24.743
10. Median Rooms/Owner HU	0.065	0.002	31.306
13. Area Income Fraction Owners	0.110	0.002	47.842
12. Area Median Income	0.101	0.002	48.726
1. Gross HH Density	0.257	0.005	52.767
9. Fraction Single Detached HU	-0.170	0.003	-54.919
15. HH Size Owner	0.164	0.002	73.577
21. Autos/HH Owners	-0.225	0.003	-84.137
8. Fraction Rental Units	-0.242	0.003	-87.131
Transit %J2W Renters			
3. Employment Access	0.019	0.003	5.546
2. Block Density	-0.038	0.003	-14.092
6. Local Job Density	-0.040	0.003	-14.300
11. Median Rooms/Renter HU	0.035	0.002	17.684
7. Local Retail Job Density	0.036	0.002	18.717
4. Retail Employment Access	-0.070	0.003	-22.529
14. Area Income Fraction Renters	0.047	0.002	28.007
9. Fraction Single Detached HU	-0.095	0.002	-45.263
16. HH Size Renters	0.083	0.002	45.324
1. Gross HH Density	0.333	0.003	98.192
22. Autos/HH Renters	-0.190	0.002	-98.886
23. Transit %J2W Owners	0.427	0.004	99.733

All endogenous variables are shaded; left-hand side variables for each nested equation are also bolded.

R-Square values:	
Autos/HH Owners	0.487
Autos/HH Renters	0.425

Gross Rent	0.546
SMOC	0.521
explTransit %J2W Owners	0.433
Transit %J2W renters	0.609

See Appendix C: for a path diagram that illustrates these coefficients. Table 6 enumerates the nature and strength of the salient relationships between the model's endogenous variables.

Table 6: Relationships of the Endogenous Variables

Endogenous Variable 1	Endogenous Variable 2	Value of Coefficient (for transformed variables)	Trends
Gross Rent	SMOC	0.372 +/- 0.002	As home ownership costs go up, rents increase.
Autos/HH Owners	SMOC	0.128 +/- 0.003	As home ownership costs go up, auto ownership increases.
Autos/HH Renters	Gross Rent	0.220 +/- 0.003	As rents goes up, auto ownership increase for renters.
Transit %J2W Owners	Autos/HH Owners	-0.225 +/- 0.003	As auto ownership goes up, transit ridership decreases for home owners.
Transit %J2W Owners	Transit %J2W Renters	0.161 +/- 0.007	As more owners use transit, more renters do as well.
Transit %J2W Renters	Autos/HH Renters	-0.190 +/- 0.002	As auto ownership goes up, transit ridership decreases for renters.
Transit %J2W Renters	Transit %J2W Owners	0.427 +/- 0.004	As more renters use transit, more owners do as well.

B. Evaluation Metrics

The complexity of SEMs has resulted in a range of metrics to assess the model goodness of fit. For the particular SEM, recommendations from R.B. Kline's *Principles and Practice of Structural Equation Modeling*, the standard text for SEMs, were followed emphasizing three metrics:

1. **Root Mean Square Error of Approximation (RMSEA)**: This metric measures error of approximation while accounting for sample size. It is an estimate of the discrepancy between

the model and the data compensating for degrees of freedom. Kline recommends the following rule of thumb: “RMSEA ≤ 0.05 indicates close approximate fit, values between 0.05 and 0.08 suggest reasonable error of approximation, and RMSEA ≥ 0.10 suggests poor fit.” A 90% confidence interval is commonly used to assess the range of the RMSEA score. The model has an RMSEA of 0.083 whose 90% confidence interval ranges from 0.082 to 0.084.

2. **Comparative Fit Index (CFI)**: This index measures the improvement in fit compared to a baseline model that assumes no population covariances for the observed variables. It analyzes the model fit examining the discrepancy between the data and the hypothesized model, while adjusting for the issues of sample size inherent in the chi-squared test of model fit. CFI pays a penalty of one for every parameter estimated. Kline suggests that CFI “values greater than roughly 0.90 may indicate reasonably good fit of the researcher’s model.” The model has a CFI of 0.925.
3. **Standardized Root Mean Square Residual (SRMR)**: This metric compares residuals between the observed and predicted variable correlations. It is the square root of the discrepancy between the sample covariance matrix and the model covariance matrix. Unlike CFI method, SRMR has no penalty for model complexity. Kline’s rule of thumb: “values of the SRMR less than 0.10 are generally considered favorable.” The model has an SRMR of 0.022.

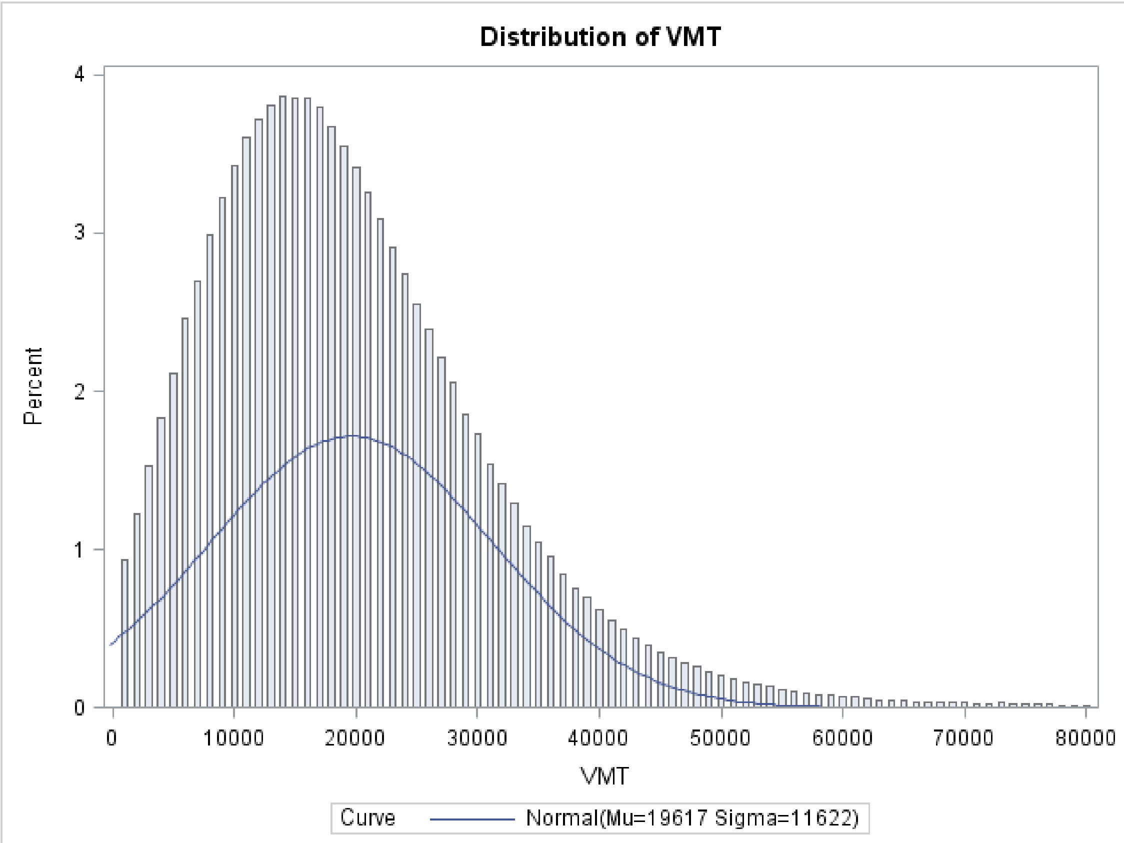
The LAIM Version 2.1 SEM meets all three of these goodness-of-fit standards, indicating that it is a good statistical model.

II. OLS Regression for Vehicle Miles Traveled

A. Independent Variable Data

As noted previously, auto use or VMT is not included in the SEM due to data limitation and is instead modeled using OLS regression. The regression model was fit using data on the total number of miles that households drive their autos, calculated from odometer readings from the Chicago and St. Louis metro areas for 2010 through 2012, obtained from the Illinois Environmental Protection Agency. Two odometer readings—for 2010 and 2012—were matched for 1,444,969 vehicles using vehicle identification numbers (VIN) to obtain data for VMT during that period. Vehicles with missing, negative and zero values were removed and frequency distributions were generated for the remaining 1,381,194 observations. The extreme values were determined to include the standard statistical approach (Howell, 1998) of mean plus 3 standard deviations (21,141 + 3*19,719 = 80,298). The histogram for the distribution of VMT data is shown in Figure 3.

Figure 3: Histogram for VMT Data for All Block Groups in Illinois metropolitan areas



The geographic area that the data covers includes a full range of place types—from rural to large city—which provides excellent fodder for calibrating a model. In order to assess the validity of this data set for the entire country, national driving records were obtained from the National Household Travel Survey (NHTS) and assigned to Census block groups using ZIP and ZCTA geographical identifications. The resulting analysis showed that the ratio of the average VMT predicted by the LAI VMT model to the average ANNMILES (the modeled value of the NHTS field ANNMILES, which is the self-reported miles driven for each auto) by Census region was 1.06,[5] suggesting that the LAI VMT model slightly underestimates auto usage nationwide. This discrepancy was expected and previous analysis suggests that it is primarily due to the fact that the vehicles represented in the Illinois EPA data were all five years of age or older, and in the aggregate older cars are driven less than newer ones. To compensate, the final value of VMT includes an adjustment factor of six percent.

B. Regression Technique

VMT is predicted using OLS regression analysis with a second-order flexible functional form. This flexible form takes into consideration all the independent variables as well as the interaction between them; i.e.,

[5] Data were averaged across each Census region (i.e. Midwest, Northeast, South, and West) due to the relatively small sample size of the NHTS.

household density, household income, and the product of household density and household income are all used as inputs. The independent variables used in the regression are essentially the same as the exogenous variables for SEM and were linearized in the same way as in the SEM analysis, with the exception that this model was run once for each household profile irrespective of tenure, using overall average income, household size and commuters per household rather than two tenure-specific versions of each variable. Since Odometer data is available only from the State of Illinois, VMT data is extrapolated to all the other states using NHTS data and modeled using OLS regression covering all Census block groups

Table 7 summarizes the independent variables used in the VMT regression. The "Number of Times Used in Combination" column indicates the number of times each variable is statistically significant and non-collinear for either the term itself, the square of the term, and/or an interaction term with another independent variable. Note that the variables highlighted in light grey were not used in this regression because they were either statistically insignificant and/or highly collinear with the other variables.

Table 8 (next page) reports the entire set of cross terms used in the models with their coefficients and values can be found in. Results from linear regression analysis shows that there is no statistical significant relationship between median rooms per housing unit and VMT, so the VMT OLS model was only run once per household type for both owners and renters together.

Table 7: Independent Variables Used in VMT Regression

Variable Name	Linear Transformation	Linearized Variable Name	Number of Times Used in Combination
Area Income Fraction	Square Root	area_income_frac	0
Area Median Income	Natural Log	area_median_hh_income	0
Median Journey to Work Miles	Natural Log	avg_d	1
Avg HH Size	Natural Log	avg_hh_size	0
Block Density	Square Root	block_density	3
Commuters/HH	None	commuters_per_hh	1
Employment Access	Natural Log	emp_gravity	0
Fraction Rental Units	Square Root	frac_renters	1
Gross HH Density	Natural Log	gross_hh_density	1
Local Job Density	Natural Log	le_jobs_total_per_acre	1
Local Retail Jobs per acre	Square Root	le_job_type_07_per_acre	3
Median Room/HU	None	median_number_rooms	0
Fraction Single Detached HU	None	pct_hu_1_detached	1

Retail Gravity	Natural Log	retail_gravity	0

Table 8: Regression Coefficients for VMT Model

Variable	Value	Standard Error	VIF
Intercept	12157.951	204.166	0.000
avg_d*commuters_per_hh	66.531	28.951	1.282
block_density^2	2577.148	400.579	3.403
block_density*gross_hh_density	-1217.119	118.619	4.708
frac_renters* pct_hu_1_detached	4.072	1.915	1.318
le_jobs_total_per_acre* le_jobs_type_07_per_acre	-34.462	8.04	4.835
le_jobs_type_07_per_acre^2	11.213	5.611	4.278

C. Evaluation Metrics

Various diagnostic methods were used to test for normality, homoscedasticity (constant variance) and autocorrelation.

1. A normality test was conducted by using correlation test for normality. In this method, we calculate both the actual residuals (ei) and expected value of the residuals under normality (Ei). The correlation coefficient between ei and Ei is calculated and compared with the table values to check for normality.
2. The Breusch Pagan test was used to test for homoscedasticity. This test is based on the residuals of the fitted model. If the test shows that there is heteroskedasticity, the regression is corrected using robust standard errors.
3. The Durbin Watson test was used to test for the presence of autocorrelation in the residuals. If this test indicates the presence of auto correlation, this is corrected using a Cochrane-Orcutt estimation.
4. Due to the inherent spatial autocorrelation for the dependent variables, a robust variance calculation was employed to estimate the statistical significance of the regression coefficients. The method for estimating the error on the coefficients tests geographical clustering at three levels: county, state, and CBSA. The testing showed that the errors estimate increased (as expected) when using this robust approach, and that the state clustering increased the error estimate the least, with the county and CBSA clustering having similar estimates; therefore the county clustering was employed.

There is a high probability that some of the independent variables are multi-collinear. To mitigate as much of this as possible, the variance inflation factor (VIF)[6] was examined. VIFs greater than 10 indicate excessive multi-collinearity that will result in unstable estimates of the regression coefficients. In this model, after eliminating coefficients with high p-value, the VIF was required to be less than 5. Values for this analysis tended to be greater than 10 to begin with, and drop perceptibly as highly multi-collinear coefficients were excluded.

Using the LAIM to Generate the Location Affordability Index (LAI)

I. Modeling Transportation Behaviors and Housing Costs

To isolate the built environment's influence on the balance between transportation and housing costs, the exogenous household variables (income, household size, and commuters per household) are set at fixed values (i.e., the "household profiles") in the Model's outputs to control for any variation they might cause. By establishing and running the model for a "household profile," any variation observed in housing and transportation costs can be attributed to aspects of the built environment (including location within the metropolitan area), rather than household characteristics.

The model was run for the eight household types in the LAI, each characterized by income, household size, and number of commuters (the same built environment inputs were used each time). These household profiles are enumerated in 9. They are not intended to match the characteristics of any particular family. Rather, they were selected to meet the needs of a variety of users, including consumers, planning agencies, real estate professionals, and housing counselors. The incomes used for seven of the eight household types are based on the median household income for each Core Based Statistical Area (CBSA) or non-metropolitan county, making the results regionally specific. The model was run for both owner and renter tenure for each profile.

Table 9: LAI Household Profiles

Household Type	Income	Size	Number of Commuters
Median-Income Family	MHHI	4	2
Very Low-Income Individual	National poverty line	1	1
Working Individual	50% of MHHI	1	1
Single Professional	135% of MHHI	1	1
Retired Couple	80% of MHHI	2	0
Single-Parent Family	50% of MHHI	3	1

[6] For a definition of VIF see http://en.wikipedia.org/wiki/Variance_inflation_factor .

Moderate-Income Family	80% of MHHI	3	1
Dual-Professional Family	150% of MHHI	4	2

MHHI = Median household income for a given area (CBSA or non-metropolitan county).

The following steps were used to run the SEM model for each household profile:

1. The model was run twice for each household profile: once for both owners and renters. This was done by using the database values for each block group for all the variables that apply to the other tenure (i.e., renters when running owner household, and owners when running renter households – see **Error! Reference source not found.**10).
2. The VMT model was run for each household type, irrespective of tenure.
3. Values for modeled Median SMOC and Median Gross Rent was evaluated and adjusted to limit outliers as follows: if the modeled value was less than the 10th percentile, overwrite the modeled value with the 10th percentile value; if over the 90 percentile, overwrite modeled value with the 90th percentile value.
4. Calculate the transportation cost, for each household type and tenure, using the unit costs developed for LAIM Version 1, but multiply by an inflation factor to determine 2014 dollars from the 2010 calculations.

These operations result in estimates of household housing costs and household transportation behaviors (autos/HH, annual VMT, annual transit trips) for both owners and renters in every block group matching each of the eight household profiles. Housing affordability can then be calculated for both owners and renters by dividing housing costs by the corresponding income for each household profile.

Table 10: Household Variables used in SEM

Modeled Variables	Owner Household Variables[7]	Renter Household Variables[8]
• Autos/HH Owners • SMOC • Transit %J2W Owners	Values from Table 9	Values from renter households in block group
• Autos/HH Renters • Gross Rent • Transit %J2W Renters	Values from owner households in block group	Values from Table 9

II. Transportation Cost Calculations

As discussed, LAIM Version 2.1 estimates three components of travel behavior: auto ownership, auto use, and transit use. To calculate total transportation costs, each of these modeled outputs is multiplied by a cost per unit (e.g., cost per mile) and then summed to provide average values for each block group.

---

[7] Household Income Owners, Household Size Owners, and Commuters per Household Owners
[8] Household Income Renters, Household Size Renters, and Commuters per Household Renters

This operation is performed for the transportation behavior estimates generated for each of the eight household types by tenure.

A. Auto Ownership and Auto Use Costs

The Consumer Expenditure Survey (CES) from the U.S. Bureau of Labor Statistics is the basis for the auto ownership and auto use cost components of the LAI. Research conducted by Diane Schanzenbach, PhD and Leslie McGranahan PhD[9], which included a range of new and used autos, examined expenditures based on the 2005-2010 waves of the CES. This identified a path to overcome the limitations of other measures that focused primarily on autos less than five years old. Based on the research, expenditures are represented in inflation-adjusted 2010 dollars using the Consumer Price Index for all Urban Consumers (CPI-U).[10] Expenses were analyzed for households in each of five income bands ($0-$19,999; $20,000-$39,999; $40,000-$59,999; $60,000-$99,999; and, $100,000 and above) and multiplied by modeled autos per household and annual VMT for the appropriate income range. LAI Version 2.1 uses an inflation factor of 1.6[11] to adjust to 2014 dollars.

Expenditures related to the purchase and operation of cars and trucks are divided into five categories:

- Average annual service flow value[12] from the time the vehicle was purchased to the time the consumer responded to the CES;
- Average annual finance charge paid;
- Ownership Costs: cost of continuing to own a purchased vehicle even if it is not driven;
- Drivability Costs: cost of keeping the vehicle in drivable shape, e.g. maintenance and repairs; and
- Driving Costs: cost of the fuel used to drive the vehicle.

[9] http://www.locationaffordability.info/downloads/Auto%20Cost.pdf
[10] For LAI version 2.1 these figures are adjusted to 2014 dollars.
[11] http://www.bls.gov/data/inflation_calculator.htm
[12] Service flow is the average annual dollar amount of depreciation the vehicle has lost over the time of ownership.

Table 11: Per-Vehicle Costs by Income Group among Households with at Least One Vehicle

Income group number and range	Average Annual Service Flow (1)	Finance Charges (2)	Per vehicle (fixed) ownership costs (3)	Per vehicle (variable) drivability costs (4)	Per vehicle fuel costs (5)	Number of vehicles (6)	Average Ratio drivability to fuel costs (7)
1 ($0-$19,999)	$2,396	$73	$657.3	$400.8	$1,182.0	1.4	0.34
2 ($20,000-$39,999)	$2,478	$133	$732.0	$421.1	$1,369.5	1.6	0.31
3 ($40,000-$59,999)	$2,586	$182	$755.6	$458.8	$1,494.2	1.9	0.31
4 ($60,000-$99,999)	$2,727	$211	$758.6	$477.6	$1,552.8	2.2	0.31
5 ($100,000 & above)	$3,139	$201	$836.6	$593.1	$1,635.6	2.5	0.36
Overall average	$2,717	$165	$752.5	$474.5	$1,460.9	1.9	0.32

The general formula for calculating of auto costs is:

$$Cost = A * \left(V_{sf} + V_{fc} + V_{fixed}\right) + \left(\frac{VMT}{MPG}\right) * G * (1 + R)$$

Where

A = Modeled autos per household

V_{sf} = Per vehicle service flow cost from **Error! Reference source not found.** 11 (1) – for the a ppropriate income group

V_{fc} = Per vehicle finance charge from **Error! Reference source not found.** 11 (2) – for the a ppropriate income group

V_{fixed} = Per vehicle (fixed) ownership cost from **Error! Reference source not found.** 11 (3) – for t he appropriate income group

VMT = the modeled annual household VMT

MPG = the national average fuel efficiency (20.7 mpg for 2008)

G = the cost of gas per gallon (average annual regional cost for 2012)[13]

R = the Average Ratio drivability to fuel cost from **Error! Reference source not found.** 11 (7) – for t he appropriate income group

[13] U.S. Department of Energy, Energy Information Administration. "Petroleum & Other Liquids." Accessed from http://www.eia.gov/petroleum/gasdiesel/.

B. Transit Use Costs

Transit cost data were obtained from the 2012 NTD.[14] Specifically, we looked at directly operated and purchased transportation revenue as reported by each transit agency in the database.[15] Most transit agencies serve only one CBSA, but there are a number of larger systems that serve multiple CBSAs, which requires their revenue be allocated proportionately among the CBSAs covered. This allocation was based on the percentage of each transit agency's bus and rail stations within each CBSA, and how much service is provided at each stop.

By way of illustration, consider a hypothetical transit agency serves two CBSAs and has a total of 1000 bus stops, 850 of which are located in the primary CBSA ($CBSA_1$) and 150 stops extend into a neighboring CBSA ($CBSA_2$). A simple approach would be to allocate 85 percent of the transit revenue to $CBSA_1$ and the remaining 15 percent to neighboring $CBSA_2$. However, this simple allocation does not take into account the frequency of service at each stop. To account for service frequency, if each bus station in $CBSA_1$ is served by a bus 1000 time a week (about a bus every 10 minutes) and bus stations in $CBSA_2$ are served 200 time a week (a little more than once an hour), the fraction of the revenue for $CBSA_1$ would be closer to:

$$CBSA_1 = \frac{1000 * 1000}{(1000 * 1000) + (200 * 85)} = 98\%$$

which would leave $CBSA_2$ with only 2 percent. Neither of these allocation methods is perfect; for instance, it is likely that low frequency buses from another CBSA would have higher revenue per trip, in which case this method would underestimate $CBSA_2$'s revenue. In order to minimize this discrepancy, the LAIM allocates revenue from each transit agency using the weighted average of the two methods.

To estimate average household transit costs, each metropolitan area's estimated total transit revenue is allocated to block groups based on the modeled value of the percentage of transit commuters and the total households within each block group. This is done by calculating the number of transit commuters for each block group, summing across block groups to estimate the total number of transit commuters in the metropolitan area, and then allocating the metro-wide transit revenue to block groups according to the proportion of the region's commuters living in each. The average household transit cost for each block group is then derived by dividing that block group's allocation of transit revenue by number of households.

This same method of allocating regional transit revenues to block groups is used for allocating transit trips. Using the overall unlinked trip numbers also reported to the NTD, the average number of household transit trips for each block group is estimated by finding the total number of annual trips in each metropolitan area and allocating them proportionally to block groups based on number of

[14] https://www.transit.dot.gov/ntd/data-product/2012-table-26-fare-passenger-and-recovery-ratio
[15] Demand response revenue is not factored into this analysis.

households and the percent of journey to work trips.[16] The following summarizes all of the required calculations:

1. Total transit commuters (metropolitan area)
    a. Commuters in each block group = [number of households] X [commuters per household]
    b. Transit commuters in each block group = [number of commuters] X [estimated percentage of commuters using transit]
    c. Transit commuters in each metropolitan area = sum of transit commuters in all block groups in the metropolitan area
2. Annual transit trips and transit fares paid (metropolitan area) – Transit trips and transit revenue are reported to the National Transit Database by transit agency, not CBSA. As a result, for larger transit systems that serve multiple CBSAs, we need to allocate transit agency trips and farebox revenue (which should be equal to the total fares paid by transit riders) to CBSAs according to the proportion of the transit agency's coverage area and frequency of service in each one. For instance, the Massachusetts Bay Transportation Authority (MBTA) commuter rail serves stops in the Boston, Worcester, and Providence CBSAs, so the commuter rail's reported farebox revenues and trips are allocated to those three regions according to the combination of proportion of total stops and service frequency in each.
    a. Annual transit trips (metro area)
    b. Annual transit fares (metro area)
3. Annual transit trips and transit fares per transit commuter (metropolitan area)
    a. Annual transit trips per transit commuter = [2a] / [1]
    b. Annual transit fares per transit commuter = [2b] / [1]
4. Estimated transit commuters per household (each household profile @ block group) = [commuters per household] X [estimated percentage of commuters using transit]
5. Estimated transit trips and transit fares per household (each household profile @ block group)
    a. Estimated annual transit trips per household = [4] X [3a]
    b. Estimated annual transit fares per household = [4] X [3b]

There are a number of metropolitan areas without sufficient information on transit stop locations and/or no revenue listed in the NTD. The average from the allocation calculation described in the previous paragraph is used for these metropolitan areas. The average transit costs are then allocated to the block

[16] This normalization method assumes that the transit use for the journey to work is a good surrogate for overall transit use.

group level based on the percentage of transit commutes and household commuter counts. The end result is an average household transit cost at the block group level.

Appendix A: LAIM Version 2 Development

During beta testing of the LAP Version 1 and subsequent discussions prior to the site's public launch, a number of reviewers suggested that the LAIM Version 1 could potentially be enhanced if the model was able to account for interaction effects.

Many advances in statistics have enabled the creation of more nuanced and sophisticated models for explaining urban phenomena along these lines. One approach that has gained currency in urban planning studies is a simultaneous (or structural) equation model (SEM). For a set of related OLS models, an SEM approach allows the dependent (left-side) variables for one or more regression equations to be included as independent (right-side) variables in other regression equations if these other independent variables could be expected to impact that equation's output. The completeness of the SEM model requires that the number of equations equal the number of endogenous variables. In the current method, six endogenous variables and six equations were used to model auto ownership, housing costs and transit usage.

This approach has clear utility for the LAI Model, which uses a specific set of independent variables describing the built environment and demographics to predict a number of interrelated transportation behaviors and housing costs. SEM better incorporates and accounts for interaction effects on the model's dependent variables, resulting in a model that has greater econometric validity.

The development process for LAIM Version 2 was highly iterative: many proposed models were tested and discarded for a variety of reasons, but each estimated model provided information. The final model used for LAIM Version 2, like all models, is not a perfect representation of reality. However, it is the best attempt to balance two competing goals: an explanatory model that highlights key interactions between variables, and a predictive model that can be employed to power the website data tools. Given these two goals, improved predictivity was to some extent prioritized at the expense of parsimony.

The final SEM includes endogenous variables housing costs, automobile ownership, and transit usage for both homeowners and renters as well as 18 exogenous variables. Auto use or annual vehicle miles traveled (VMT) continues to be modeled using OLS because VMT data is only available from the State of Illinois, and it does not distinguish between auto owners who rent versus those who own their home.

I. Advances over LAIM Version 1

A. Model Refinements

The use of the SEM, as well as additional development work, led to two innovations in the model structure as enumerated below.

1. **Model Integration**: The power of the SEM was leveraged to reduce the number of necessary models. SEM model is very flexible and handles not only with multiple linear regressions, but with a system of regression equations. In contrast to ordinary least square regression, SEM

considers several equations simultaneously. The new model structure allows a single model to predict housing costs, auto ownership levels, and transit commute mode shares rather than having separate equations for each (although VMT continues to be modeled separately). This is the inherent benefit of the SEM.

2. **Model Comprehensiveness**: The combination of the SEM approach and the refined variables allowed development of a single model for the entire nation rather than separate models for urban and rural areas. This was achieved by focusing on county level data rather than CBSA data for rural counties and taking advantage of the feedback inherent in SEM to use the share of transit commuters as a proxy variable for transit service levels. Previously, the model was split between areas where transit service levels were known and areas where transit service levels were unknown. SEM allows transit mode share to be simultaneously an explanatory and a response variable. The reduction in the number of input (exogenous) variables reduces the goodness of fit for the places where explicit transit supply data was available, but enhances the simplicity of the model, making it possible to develop only one model for all census block groups (both urban and rural) for the entire country.

B. Variable Refinements

During the development of LAIM Version 2, the original set of variables was reconsidered and refined as possible. A short description of these refinements follows.

1. **Local Amenities**: Local job measures were developed as a proxy for local amenities. This information is helpful in determining whether one could live in an area without a car and still have access to basic needs, such as shopping.
2. **Income Scaling:** A variable that scales income based on the regional median income within Core Based Statistical Areas (CBSAs) and the county median income in rural areas outside of a CBSA. This adjustment improves the ability to offer an "apples-to-apples" comparison of purchasing power, particularly for auto-ownership decisions. It is also the relevant median income within the model to appropriately estimate housing expenses based on the local market. This "mixed" approach, using the regional median for CBSAs and the county median for rural areas, fits the data better than a simple CBSA or county-based approach.
3. **Housing Characteristics**: Housing stock data, specifically percent of single-family detached housing units and the number of rooms per dwelling unit by occupied tenure, were incorporated into the model.
4. **Tenure Split**: Population data was split based on whether the respondents own or rent their residence. This affects variables tied to people (household size, income, transit mode shares, etc.), but not those tied to the surrounding environment (household density, job density, etc.). The resulting model structure provides added insight into the decisions of renters and owners although it reduces the predictive power of the overall model by a few percentage points. However, given the strong theoretical justification for considering renters and owners separately, it was decided to include this split in the final model.

II. Model Specification

A. Endogenous Variable Interactions

The first step in developing an SEM is to develop the model specification, using a set of hypotheses that illustrate the relationship between the various input variables. The endogenous variables (below) are each predicted by individual regression models nested within the SEM and are all interrelated:

- Autos/Household Owners
- Autos/Household Renters
- Gross Rent
- Selected Monthly Ownership Costs (SMOC)
- Transit Percent Journey to Work (%J2W) Owners
- Transit %J2W Renters

(following page) is a schematic representation of the relationships that drove the decision to add feedback in the SEM between the endogenous variables. In principle, causality can go both ways; in the actual implementation, it was found that once causality is explained in one direction, the other direction is either not statistically significant or markedly less significant, and the goodness of fit is reduced. For example, having SMOC in the homeowner auto ownership equation obviates the need for putting homeowner auto ownership into the SMOC equation. The one exception to this is the interaction between owner and renter transit use; in these cases, both interactions were found to be important and thus were included in the final model (denoted by the double headed arrow).

2, following the schematic, shows the hypothesis and the relationships in the final model. Interactions are limited to only those of the same tenure, unless the endogenous variables are of the same behavior (i.e., Auto Use by Owners interacts with Auto Ownership by Renters but not with Transit %J2W Renters or Gross Rent).

Figure 4: Schematic Representation of the Relationships between the Endogenous Variable Implemented in the SEM

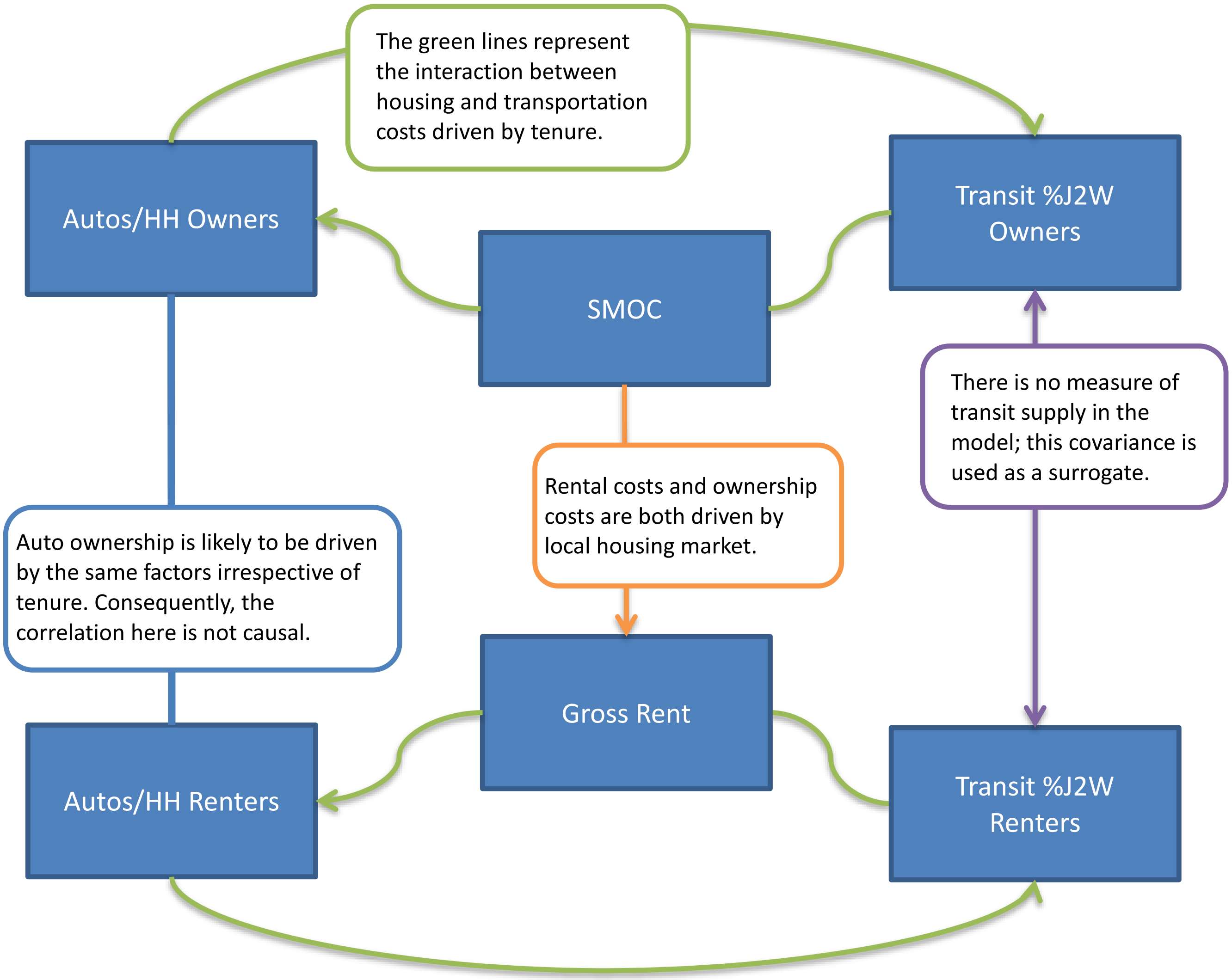


Table 12: Hypothesis of Endogenous Variable Interactions

Variable 1 (V_1)	Variable 2 (V_2)	Working Hypothesis	Interaction Used
Autos/Household Owners	Autos/Household Renters	Auto ownership is driven by the same factors independent of tenure. The correlation observed here is coincidental and not causal so no explicit connection is used in model.	None
Autos/Household Owners	SMOC	Auto ownership and housing costs are both very large components of a household's budget. Thus these two measures are totally constrained by the budget and are very dependent on one another.	One Way ($V_2 \rightarrow V_1$)
Autos/Household Owners	Transit %J2W Owners	Auto ownership and transit use are obviously related.	One Way ($V_1 \rightarrow V_2$)
Autos/Household Renters	Gross Rent	Auto ownership and housing costs are both very large components of a household's budget. Thus these two measures are total constrained by the budget and are very dependent on one another.	One Way ($V_2 \rightarrow V_1$)
Autos/Household Renters	Transit %J2W Renters	Auto ownership and transit use are obviously related.	One Way ($V_1 \rightarrow V_2$)
SMOC	Gross Rent	Local housing market conditions depend on household formation, interest rates, household net worth, labor market conditions and other fundamental factors such as housing stock. In some models, these fundamental factors determine long run equilibrium housing costs as reflected in rental costs, while short run ownership costs fluctuate around long run equilibrium (rental) values, with short run fluctuations driven in part by the inventory/sales ratio.	One Way ($V_1 \rightarrow V_2$)
SMOC	Transit %J2W Owners	Unlike the relationship between housing cost and auto ownership, the cost of transit is relatively low thus the constraint driven by a household's budget is less rigid. Thus there is no strong reason to expect an interaction and none was observed.	None

Variable 1 (V_1)	Variable 2 (V_2)	Working Hypothesis	Interaction Used
Gross Rent	Transit %J2W Renters	Unlike the relationship between housing cost and auto ownership, the cost of transit is relatively low thus the constraint driven by a household's budget is less rigid. Thus there is no strong reason to expect an interaction and none was observed.	None
Transit %J2W Owners	Transit %J2W Renters	Transit use is driven by the same factors independent of tenure. The correlation observed is driven by non-measured exogenous variables. Since this model has no transit supply or access measure, this interaction serves as a surrogate.	Two Way

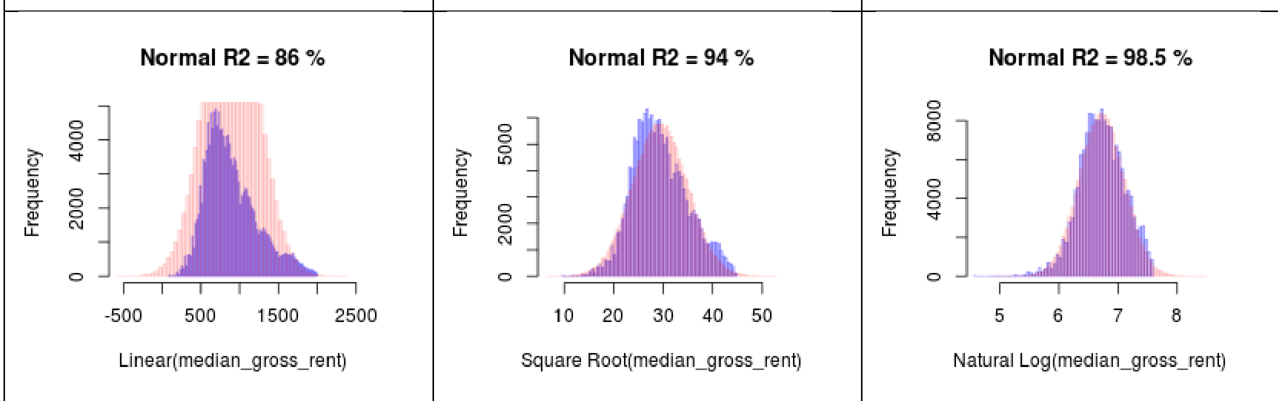
B. Variable Transformation

Once relationships between endogenous variables have been hypothesized, a preliminary model can be constructed. In LAIM Version 2, SEM variables (Table 13, next page) are transformed to allow for better fits for non-linear relationships. As shown in Figure 6 (below), a typical approach to transforming variables is used. This is the same approach as that used in the original LAIM, i.e., pick the transformation that produces the most normal distribution for each variable – both the endogenous and exogenous. The frequency distributions in Figure 5 represent an example for median gross rent. The components of the figure are described below:

- Purple bars represent ACS data
- Red bars represent a Gaussian (or normal) distribution with the same mean and standard deviation as the census data
- The "Normal R^2" value is coefficient of determination of the ACS data to the normal distribution.

Figure 5: Example of Linear Transformation

Linear (No Transformation)	Square Root - $\sqrt{x}$	Natural Log – ln(x) (used in SEM for this variable)



By evaluating the exogenous variables to observe how non-linear the relationships between them are, a transformation is chosen to reduce non-linear effects. In the SEM approach used in LAIM Version 2, the transportation endogenous variables were not transformed; however, housing costs variables (gross rent and SMOC) are transformed using the natural log as in LAIM Version 1.

The transformed variable was subtracted by the mean of the transformed variable's distribution; this difference was then scaled by one over the standard deviation of the entire distribution. The resulting variable (Z) used in the SEM analysis is:

$$Z = \frac{f - \bar{f}}{StDev}.$$

Where f is the transformed variable, $\bar{f}$ is the mean of the distribution of f and *StDev* is the standard deviation of the distribution of f.

Table 13: Variables Used to Estimate the Model, with Transformations and Descriptive Statistics

Name	Transformation	Mean of Transformed Variables	Standard Deviation of Transformed Variables
Area Income Fraction Owners	Natural Log	0.092	0.352
Area Income Fraction Renters	Natural Log	-0.569	0.448
Area Median Income	Natural Log	10.856	0.209
Median J2W Miles	Natural Log	2.286	0.673
HH Size Owner	Natural Log	0.949	0.232
HH Size Renters	Natural Log	0.891	0.332
Block Density	Square Root	0.288	0.173
Commuters/HH Owners	Linear	1.170	0.332
Commuters/HH Renters	Linear	1.018	0.358
Employment Access	Natural Log	9.251	1.436
Fraction Rental Units	Square Root	0.579	0.182
Gross HH Density	Square Root	1.380	1.180
Local Retail Jobs per acre	Square Root	0.373	0.384
Local Job Density	Square Root	1.150	1.217
Median Rooms/Owner HU	Linear	6.150	0.930
Median Rooms/Renter HU	Linear	4.649	1.036
Fraction Single Detached HU	Linear	62.152	27.683

Retail Gravity	Natural Log	7.057	1.376
Autos/HH Owners	Linear	1.949	0.421
Autos/HH Renters	Linear	1.353	0.492
Gross Rent	Natural Log	6.727	0.386
SMOC	Natural Log	7.231	0.391
Transit %J2W Owners	Linear	3.814	10.053
Transit %J2W renters	Linear	6.012	14.008

J2W = Journey to Work HH = Households
HU = Housing Units SMOC = Selected Monthly Ownership Costs
Endogenous variables are shaded.

This standardization—converting to z-scores—was applied to each variable to enable the SEM function in R[17] to handle the wide variation in values. However, it has the added benefit of making the model more transparent in two ways: 1) there is no need for an intercept in the regression equation, and 2) the coefficients are equal to the magnitude of the change expected in the transformed endogenous variable when the transformed exogenous variable is increased or decreased by one standard deviation. Furthermore, this standardization does not change the shape of the original distribution.

C. Variable Selection

Table 14 lists the variables used in the original LAIM but not included in LAIM Version 2.

Table 14: LAIM Version 1 Variables Dropped in LAIM Version 2

Dropped Variable	Description	Reason for Dropping
Residential Density	# of households in residential blocks	Highly correlated with gross density. Gross density can be obtained annually from the ACS rather than relying on decennial census for Residential Density.
Intersection Density	# of intersections / total land area	Encapsulated by other measures of local walkability/density (See section on B. Street Connectivity and Walkability)
Transit Connectivity Index	Transit access as a function of transit service frequency and proximity to	Replaced by transit commute share, a measure available for the

[17] R is a software programming language used for statistical analysis.

Dropped Variable	Description	Reason for Dropping
	transit nodes, weighted by observed journey to work data	entire country.
Transit Access	Shed Optimal accessible area by public transportation within 30 minutes and one transfer	Ibid
Transit Frequency of Service	Service frequency within a Transit Access Shed	Ibid
Job Diversity Index	Function of the correlation between employment in 20 different industry sectors and autos per household	Job diversity was determined to not be the best measure of local transit amenity; replaced with a count of actual local jobs.
Median Selected Monthly Owner Costs	Includes mortgage payments, utilities, fuel, and condominium and mobile home fees where appropriate	Median area SMOC is not a strong a predictor of regional housing markets so it was replaced with the area median income for each CBSA (or non-metropolitan county).
Median Gross Rent	Includes contract rent as well as utilities and fuel if paid by the renter	Ibid
Median household income		Replaced by scaled income (see point 2 on page ii of this document)

Variables listed in Table 15 were added to the model based on feedback from HUD staff and a literature review of rural VMT.

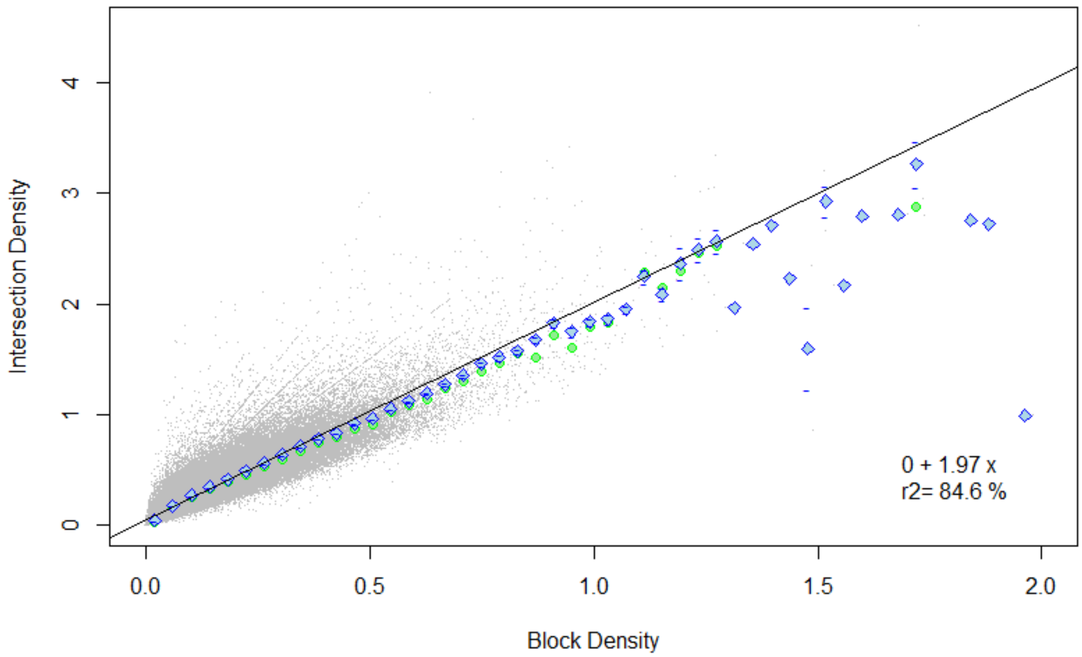
Table 15: Variables Added for the SEM/Rural Analysis

Added Variable	Description	Reason for Adding
Area Median HH Income	Median county household income for counties in Rural (non CBSA) areas, and CBSA Median household income for those within a CBSA.	To scale for regional market variations in housing cost.
Fraction Rental Units	Number of rental occupied housing units divided by all occupied housing units.	To adjust for different housing stock and use.
Local Retail Jobs per acre	Number of retail jobs within half mile of	Access to retail amenities.

Added Variable	Description	Reason for Adding
	centroid divided by land area of same.	
Local Job Density	Number of jobs within half mile of centroid divided by land area of same.	Local job access.
Median Rooms/Owner HU	Median number of rooms In housing units for owner occupied units.	Indicator of local ownership housing stock size.
Median Rooms/Renter HU	Median number of rooms In housing units for renter occupied units.	Indicator of local rental housing stock size.
Fraction Single Detached HU	Number of housing units in single family detached buildings.	Indicator of local housing type.
Retail Gravity	Same as employment gravity but only for retail jobs.	Access to regional retail amenities.
Income/Area Income Owner	Median household income divided by county median income for occupied owner housing units in Rural (non CBSA) areas and by CBSA Median income for those within a CBSA.	Scaled income (see point 2 on page ii of this document).
Income/Area Income Renter	Median household income divided by county median income for occupied owner occupied housing units in Rural (non CBSA) areas and by CBSA Median income for those within a CBSA.	Ibid

Although LAIM Version 1 used a combination of block density and intersection density, the only measure of street connectivity and walkability used in in LAIM Version 2 and beyond is block density. The addition of intersection density created a model that is slightly better in terms of prediction, but because of the very high co-linearity between these two measures, it made the model less transparent. Since block density improves the SEM model more than intersection density, block density was chosen to be included in subsequent LAIM Versions. Figure 6, which shows the correlation between the measures, illustrates just how collinear these two measures are.

Figure 6: Intersection Density (Intersections per Acre) versus Block Density (Blocks per Acre) for all U.S. Census Block Groups



D. Final Fit

The following section describes in detail final model's specification and all included variables for LAIM Version 2. The structure of the model is detailed in Table 16.

Table 16: SEM: Regression Coefficients for Endogenous and Exogenous Variables

Variables	Estimate	Std. Error	Z-Value
Autos/HH Owners			
Employment Access	-0.392	0.008	-46.738
Gross HH Density	-0.264	0.003	-84.278
HH Size Owner	0.227	0.002	107.228
Commuters/HH Owners	0.193	0.002	83.136
Fraction Single Detached HU	0.174	0.002	69.930
Area Income Fraction Owners	0.130	0.002	55.887
Fraction Rental Units	0.116	0.002	47.750
Retail Gravity	0.116	0.008	14.810
Area Median Income	0.106	0.003	40.375
SMOC	0.092	0.003	34.818
Median Rooms/Owner HU_	0.084	0.002	43.817
Block Density	-0.080	0.003	-29.157

Variables	Estimate	Std. Error	Z-Value
Autos/HH Renters			
Employment Access	-0.302	0.010	-31.343
Commuters/HH Renters	0.213	0.002	88.866
Gross HH Density	-0.200	0.003	-58.340
HH Size Renters	0.156	0.002	70.345
Area Income Fraction Renters	0.149	0.002	62.935
Gross Rent	0.142	0.003	53.859
Median Rooms/Renter HU	0.128	0.002	56.239
Fraction Single Detached HU	0.118	0.002	49.290
Retail Gravity	0.111	0.009	12.717
Area Median Income	0.085	0.003	32.581
Block Density	-0.056	0.003	-19.310
Median J2W Miles	-0.041	0.002	-18.268
Local Job Density	0.032	0.003	11.244
Gross Rent			
Retail Gravity	0.343	0.008	44.885
SMOC	0.263	0.002	117.664
Area Median Income	0.259	0.002	109.326
Area Income Fraction Renters	0.256	0.002	141.793
Median Rooms/Renter HU	0.183	0.002	92.134
HH Size Renters	0.141	0.002	78.135
Employment Access	-0.104	0.008	-12.494
Gross HH Density	0.061	0.003	21.524
Median J2W Miles	0.032	0.002	16.226
Block Density	-0.027	0.003	-10.415
Fraction Rental Units	-0.027	0.002	-12.761
SMOC			
Area Median Income	0.519	0.002	244.723
Area Income Fraction Owners	0.425	0.002	221.824
HH Size Owner	0.129	0.002	65.377
Commuters/HH Owners	-0.127	0.002	-59.079
Retail Gravity	0.113	0.007	15.221
Block Density	-0.100	0.003	-39.396
Employment Access	0.088	0.008	10.618
Fraction Single Detached HU	-0.082	0.002	-35.474
Gross HH Density	0.082	0.003	27.504
Median Rooms/Owner HU	0.081	0.002	44.880
Median J2W Miles	0.081	0.002	41.923
Fraction Rental Units	-0.037	0.002	-16.236
Local Job Density	0.030	0.002	12.305
Transit %J2W Owners			

Variables	Estimate	Std. Error	Z-Value
Gross HH Density	0.434	0.004	101.848
Transit %J2W renters	0.321	0.006	54.773
Retail Gravity	-0.251	0.008	-31.503
Autos/HH Owners	-0.215	0.002	-95.394
Employment Access	0.133	0.008	16.133
HH Size Owner	0.129	0.002	73.212
Block Density	-0.122	0.003	-48.496
Fraction Rental Units	-0.101	0.002	-45.202
Area Median Income	0.098	0.002	48.368
Fraction Single Detached HU	-0.066	0.002	-27.903
Local Retail Jobs per acre	0.050	0.002	21.345
Area Income Fraction Owners	0.029	0.002	15.900
Median Rooms/Owner HU	0.026	0.002	14.811
Transit %J2W renters			
Employment Access	0.408	0.009	46.595
Transit %J2W Owners	0.397	0.006	69.318
Retail Gravity	-0.372	0.008	-44.125
Gross HH Density	0.256	0.005	53.826
Autos/HH Renters	-0.172	0.002	-91.273
HH Size Renters	0.076	0.002	44.395
Fraction Single Detached HU	-0.072	0.002	-35.277
Local Job Density	-0.065	0.003	-21.384
Area Income Fraction Renters	0.057	0.002	28.274
Median Rooms/Renter HU	0.044	0.002	23.358
Local Retail Jobs per acre	0.040	0.003	13.299
Block Density	-0.036	0.003	-14.364

R-Square:	
Autos/HH Owners	0.550
Autos/HH Renters	0.470
Gross Rent	0.578
SMOC	0.611
Transit %J2W Owners	0.628
Transit %J2W renters	0.630

All endogenous variables are shaded; left-hand side variables for each nested equation are also bolded.

Table 17 gives the nature and strength of the salient relationships between the model's endogenous variables.

Table 17: Relationships Between Endogenous Variables

Endogenous Variable 1	Endogenous Variable 2	Value of Coefficient (for transformed variables)	Trends
Gross Rent	SMOC	0.263 +/- 0.002	As home ownership costs go up, rents increase.
Autos/HH Owners	SMOC	0.092 +/- 0.003	As home ownership costs go up, auto ownership increases.
Autos/HH Renters	Gross Rent	0.142 +/- 0.003	As rents goes up, auto ownership increase for renters.
Transit %J2W Owners	Autos/HH Owners	-0.215 +/- 0.002	As auto ownership goes up, transit ridership decreases for home owners.
Transit %J2W Owners	Transit %J2W Renters	0.321 +/- 0.006	As more owners use transit, more renters do as well.
Transit %J2W Renters	Autos/HH Renters	-0.172 +/- 0.002	As auto ownership goes up, transit ridership decreases for renters.
Transit %J2W Renters	Transit %J2W Owners	0.397 +/- 0.006	As more renters use transit, more owners do as well.

Some notable advances in LAIM Version 2:

1. **By not including residuals back into the modeled housing costs, large errors from the ACS are not reintroduced.** In LAI Version 1, once the housing costs were estimated the residual from the fit was added back into the value. A third-party review of LAI Version 1[18] suggested this measure to account for different quality of housing stock and intangibles not being modeled, but this increased the variability of the results because it included the large measurement errors from the ACS. Because new measures of housing quality have been included in the SEM model, reintroduction of the large ACS measurement error is avoided. As the SEM model used in LAI Version 2 includes variables which measure housing quality (i.e., rooms per dwelling unit, fraction of detached single family houses, and fraction of renters in the neighborhood), this source of variation is avoided. The SEM modeled values for household type 1 are overall consistent with those of LAI Version 1 (accounting for a small increase in their values) and show less variation as a result.

[18] Econsult Solutions conducted a third-party review of LAIM Version 1 to assess the validity of the model and provide recommendations for potential improvements.

2. **Different transportation costs are modeled by tenure for each of the eight household types.** The advantage of including tenure into the model is that it delivers a better estimate of transportation cost for renters versus owners.
3. **The My Transportation Cost Calculator (MTCC) now includes a progressively more accurate estimate of the users' housing and transportation costs.**[19] A new text box on each tab of the calculator takes advantage of the SEM using the progression of choices made by the user.
4. **National coverage includes rural areas.** SEM allows transit mode share to be simultaneously an explanatory and a response variable. The reduction in the number of input (exogenous) variables reduces the goodness of fit for the places where explicit transit supply data was available, but enhances the simplicity of the model, making it possible to develop only one model for all census block groups (both urban and rural) for the entire country.

[19] Note that My Transportation Cost Calculator is no longer available.

Appendix B: Scatter Plots of Endogenous Variables vs. an Example Exogenous Variable

The following plots show the relationships between some of the exogenous variables and the endogenous variables from LAI version 2. Note that in each plot there are approximately 200,000 points, depending on the data suppression in the ACS. Each plot has the following features:

- Small grey dots – values for each census block group where there is valid data (i.e. no ACS data suppression);
- Blue diamonds with blue dashed above and below – mean value of the y variable in 50 bins of the x variable, and the blue lines represent the standard error on the mean (when there is no lines this indicates that there are only one block group in this bin);
- Solid green circles – median value of the y variable in 50 bins of the x variable;
- Black line – the linear fit of the y variable with the x variable (note that for many this shows how non-linear many of these relationship are); and
- Text in lower right corner – the equation for the line and the R^2 of the linear fit.

Figure 7: Scatter plot for block group autos per household (owners) by frequency

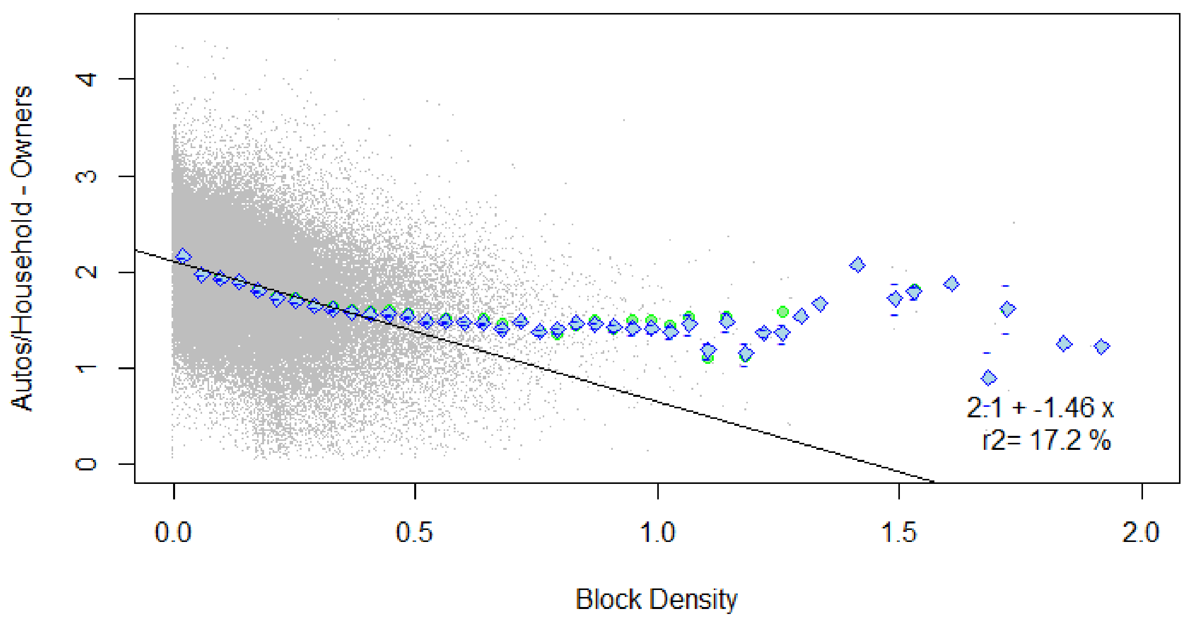


Figure 8: Scatter plot for block group autos per household (renters) by frequency

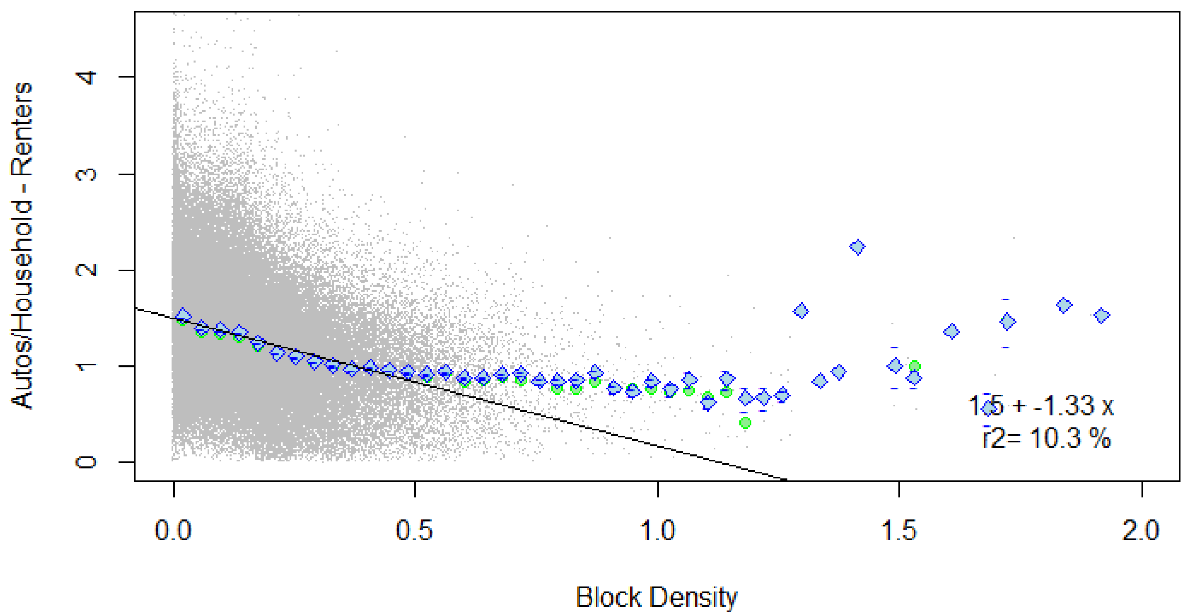


Figure 9: Scatter plot for block group for median Select Monthly Ownership Costs by frequency

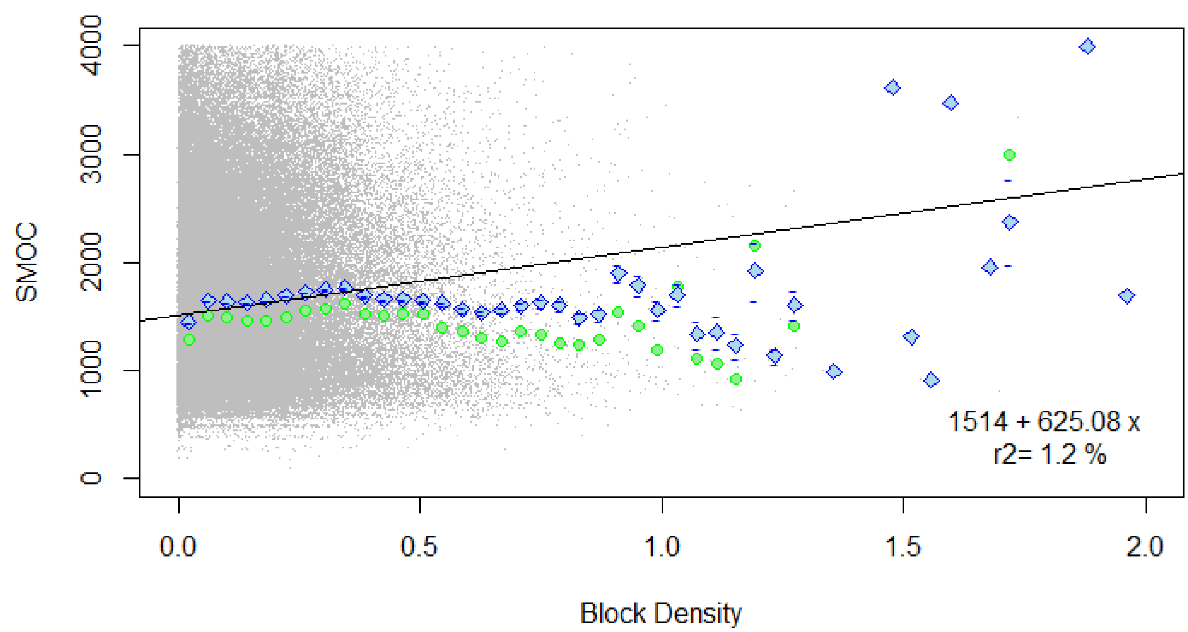


Figure 10: Scatter plot for block group median Gross Rent by frequency

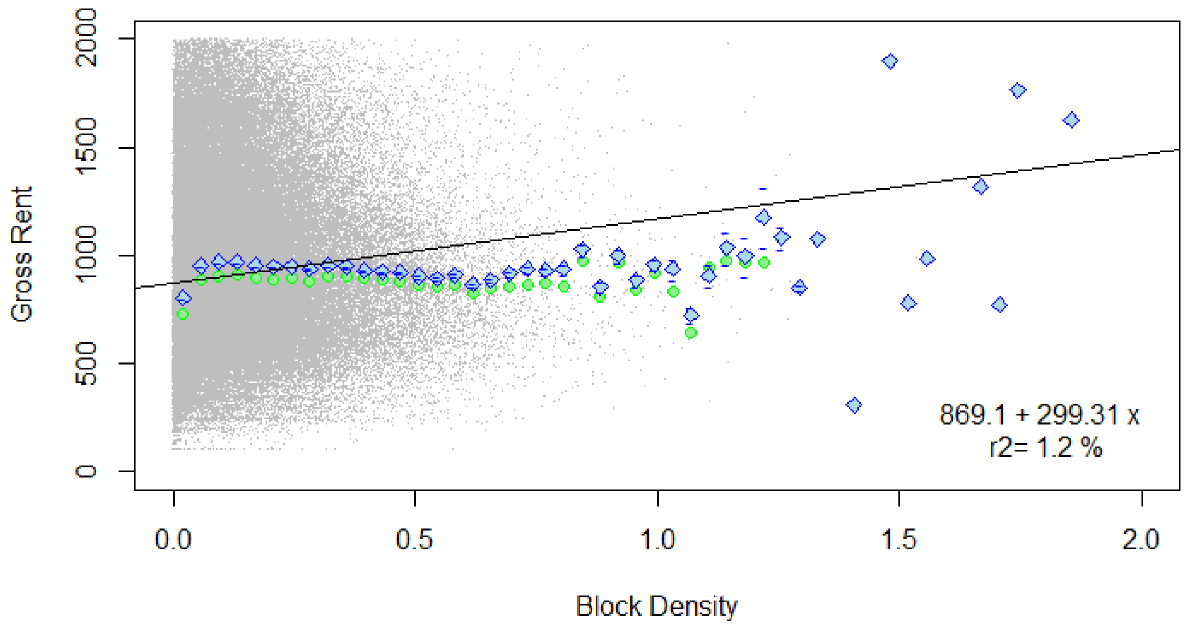


Figure 11: Scatter plot for block group Percent Journey to Work by Transit (owners) by frequency

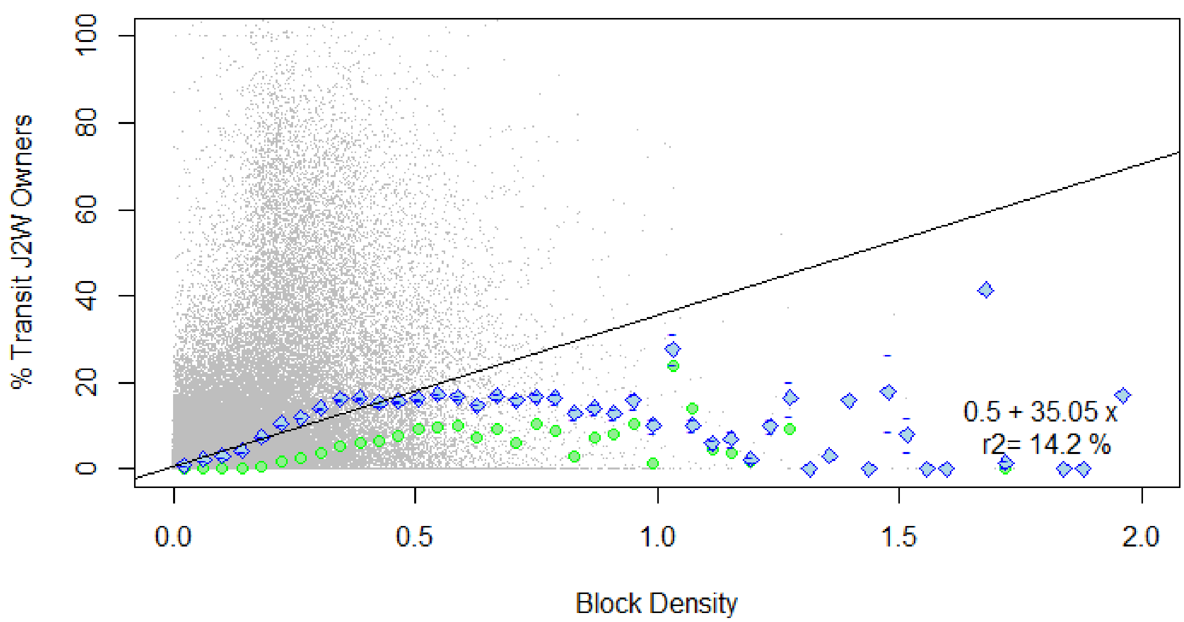
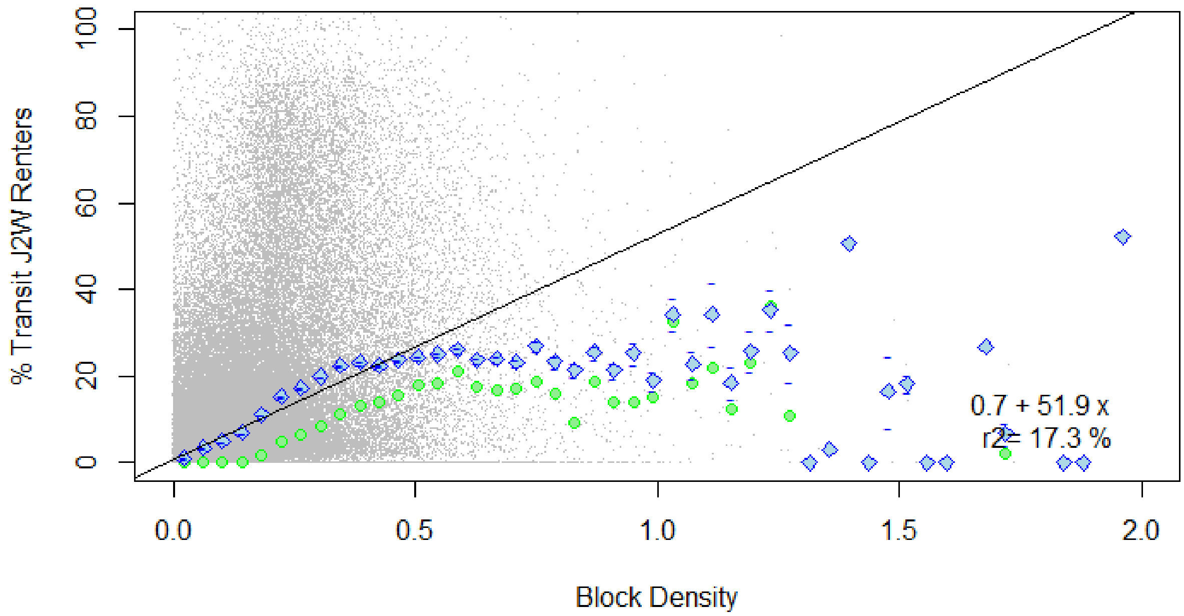


Figure 12: Scatter plot for block group Percent Journey to Work by Transit (renters) by frequency



Appendix C: Path Diagrams

Figures 13 and 14 (following pages) are different graphical representations that show the strength of the relationships between all the variables in the SEM fit. The color is either:

- Green – indicating that the relationship is positive, i.e., as Income goes up SMOC increases
- Red – indicates that the relationship is negative, i.e., as employment gravity goes up auto ownership goes down.

The width and darkness of the line indicates the strength of the relationship: wider darker lines indicate strong relationships while thinner lighter lines indicate weaker relationships. The path diagram illustrated in Figure 13 shows the values of the standardized variables used for LAIM Version 2.1; Figure 14 is the same diagram but with a different layout.

Figure 13: Path Diagram for SEM Model

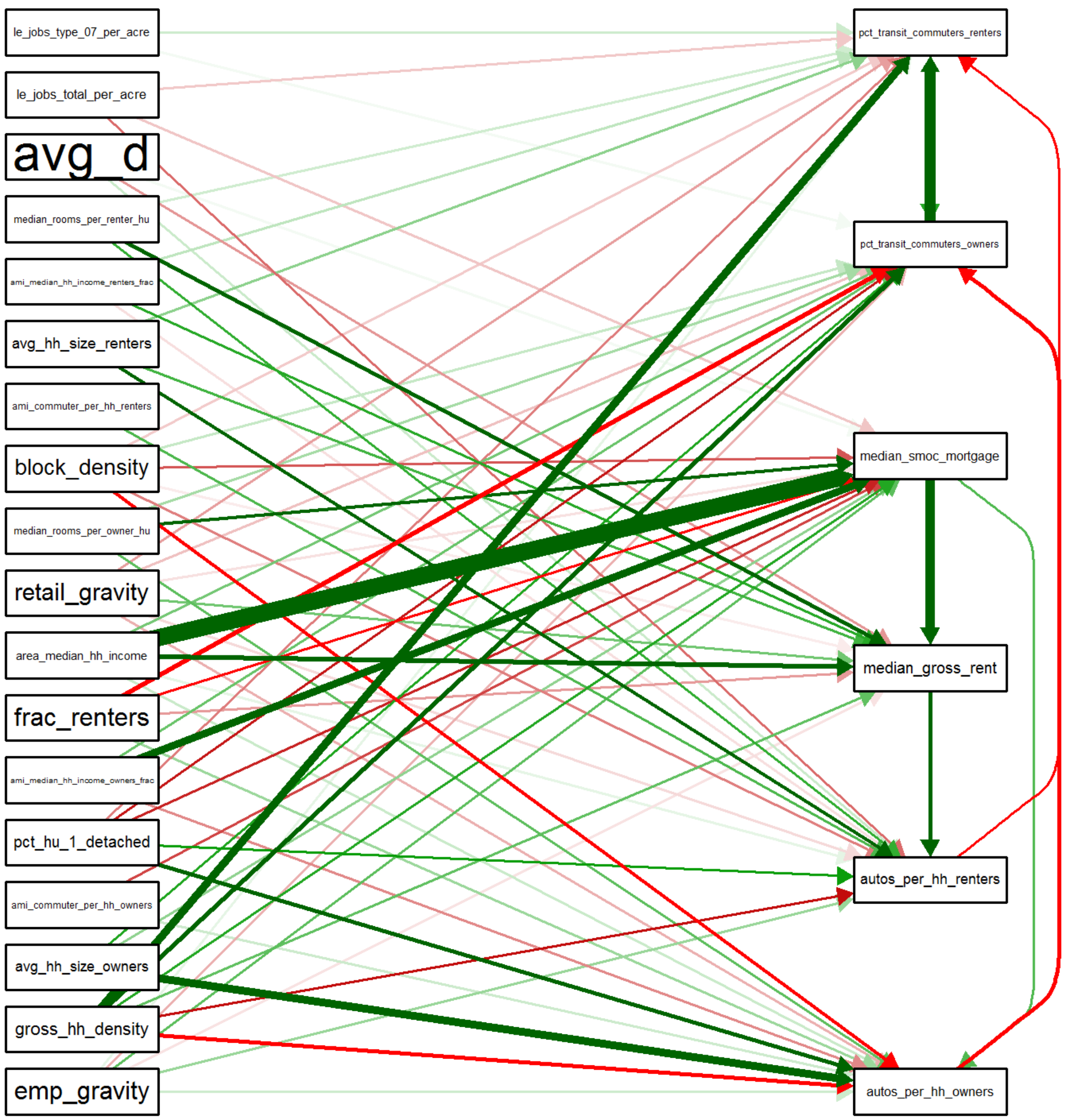
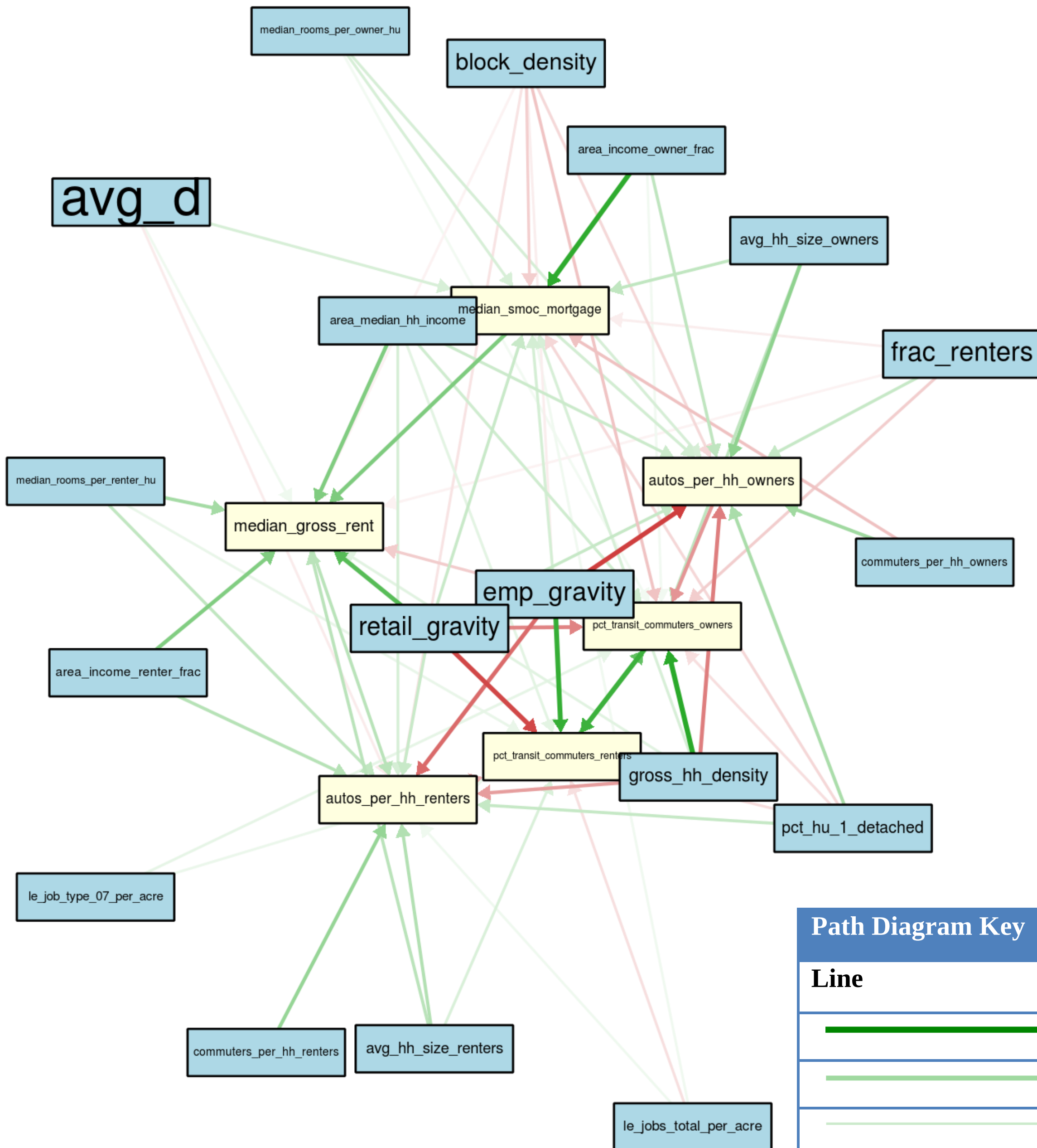


Figure 14: Path Diagram for SEM Model - Alternative Layout



Path Diagram Key	
Line	Value
	0.40
	0.20
	0.04
	-0.04
	-0.20
	-0.40

Appendix D: Simultaneous Equation Models

The standard way to write the structural model is as follows:

$$\gamma_{11}Y_1 + \gamma_{21}Y_2 + \cdots + \gamma_{M1}Y_M + \beta_{11}X_1 + \beta_{21}X_2 + \cdots + \beta_{K1}X_k + e_1 = 0$$

$$\gamma_{12}Y_1 + \gamma_{22}Y_2 + \cdots + \gamma_{M2}Y_M + \beta_{12}X_1 + \beta_{22}X_2 + \cdots + \beta_{K2}X_k + e_2 = 0$$

.

.

.

$$\gamma_{1M}Y_1 + \gamma_{2M}Y_2 + \cdots + \gamma_{MM}Y_M + \beta_{1M}X_1 + \beta_{2M}X_2 + \cdots + \beta_{KM}X_k + e_M = 0$$

Where

Y_1, Y_2,…, Y_M are T x 1 vectors containing T observations for each of the M endogenous variables;

X_1, X_2,…, X_K are T x 1 vectors containing T observations for each of the K exogenous variables;

e_1, e_2,…, e_M are T x 1 vectors containing T observations for each of the M random variables;

γ's are M^2 structural parameters

β's are MK structural parameters

The disturbance terms are assumed to have the following properties:

$E\left(e_j\right) = 0$

$cov\left(e_i, e_j\right) = \sigma_{ij}$

So the covariance matrix is represented as:

$$(ee') = \begin{pmatrix} \sigma_{11}I & \sigma_{12}I \ldots. & \sigma_{1M}I \\ \sigma_{21}I & \sigma_{22}I \ldots. & \sigma_{2M}I \\ . & . & . \\ . & . & . \\ \sigma_{M1}I & \sigma_{M2}I \ldots. & \sigma_{MM}I \end{pmatrix}$$

Since there are 6 endogenous variables (M=6) and 18 exogenous variables (K = 18) in the model, there are 36 parameters to estimate in the SEM model. The estimates from SEM model for LAIM Version 2.1 are shown in Table 5. For the SEM model, for the rank condition to be satisfied it is necessary that K > M. In the current SEM model, with 18 exogenous variables and 6 variables where K > M and this case is defined as over identified.

References

Ewing, Reid, Rolf Pendall, and Don Chen. Measuring sprawl and its transportation impacts. Transportation Research Record: Journal of the Transportation Research Board, (1831), 175-183.

Howell, D.C. (1998). Statistical methods in human sciences. New York: Wadsworth.

Kenny DA: Measuring model fit. 2016, accessed May 26, 2016, [http://davidakenny.net/cm/fit.htm]

Kline, R. B. (2015). Principles and practice of structural equation modeling. Guilford publications.

Sinnott, R. (1984). Virtues of the Haversine. Sky and telescope, 68:158.

Zhao, F., Chow, L.-F., Li, M.-T., Ubaka, I., & Gan, A. (2003). Forecasting Transit Walk Accessibility: Regression Model Alternative to Buffer Method. Transportation Research Record: Journal of the Transportation Research Board, 1835(1), 34-41.