

# **Location Affordability Index Version 3 Data and Methodology**



# Contents

I. Background .....	3
II. LAIM Version 3 .....	5
A. Basic Index Structure .....	5
B. Data Sources .....	5
C. Variables.....	6
1. Household Density .....	7
2. Street Connectivity and Walkability.....	7
3. Employment Access and Diversity .....	8
4. Housing Characteristics.....	9
5. Household Characteristics .....	10
6. Housing Costs.....	11
7. Household Transportation Behavior.....	11
D. Simultaneous Equations Model.....	11
1. Endogenous Variable Interactions.....	12
2. Variable Transformation .....	15
3. Variable Standardization.....	17
4. Final Model Structure and Formula .....	18
E. Modeling Vehicle Miles Traveled.....	22
III. Using the LAIM to Generate the Location Affordability Index (LAI).....	24
A. Modeling Transportation Behaviors and Housing Costs .....	24
B. Using Modelled Transportation Behavior to Calculate Transportation Costs .....	25
1. Auto Ownership and Auto Use Costs.....	26
2. Transit Use Costs.....	27
C. The Location Affordability Index database.....	28
Appendix 1: Scatter Plots of Endogenous Variables vs. an Example Exogenous Variable .....	32
Appendix 2: Path Diagrams .....	36
Appendix 3: Distributions of tracts by percentile of household profile income.....	39
Appendix 4: Version 2 Model Development .....	43
A. Model Refinements.....	43
B. Variable Refinements .....	44

---

*Tables*

Table 1: Changes Between Location Affordability Index Model Version 2 and 3..... 3

Table 2: Overview of LAIM Variables ..... 6

Table 3: Endogenous Variable Interactions – Hypothesized Mechanisms ..... 14

Table 4: R<sup>2</sup> for OLS fit using all variables, but changing the transformation function for x=Employment Access Index..... 15

Table 5: Linearization transformation functions from LAIM3 and LAIM2 ..... 16

Table 6: Variables Used to Estimate the Model, with Transformations and Descriptive Statistics..... 17

Table 7: Relationships of the Endogenous Variables..... 20

Table 8: Summary of R<sup>2</sup> Values for each Endogenous Variable..... 21

Table 9: Independent Variables Used in VMT Regression ..... 23

Table 10: Regression Coefficients for VMT Model in order of R<sup>2</sup> reduction..... 23

Table 11: LAI Household Profiles ..... 25

Table 12: Household Variables used in SEM..... 25

Table 13: Per-Vehicle Costs by Income Group among Households with at Least One Vehicle..... 26

Table 14: NTD Transit Agencies by Number of Service Areas..... 27

Table 15: LAI Data Dictionary: Model Inputs ..... 28

Table 16: LAI Data Dictionary: Model Outputs for Each Household Profile ..... 29

## I. Background

First launched by the U.S. Department of Housing and Urban Development (HUD) and Department of Transportation (DOT) in November 2013, the Location Affordability Index (LAI) provided ubiquitous, standardized household housing and transportation cost estimates at the Census block-group level for the majority of the populated area of the United States. These estimates were generated using the *Location Affordability Index Model* (LAIM Version 1), a combination of statistical modeling and data analysis using data from several federal sources. They were presented on the site in the form of two data tools, both of which were available on the Location Affordability Portal hosted and maintained by HUD: the Location Affordability Index (LAI), which visually represents outputs for eight different household profiles in the form of a national map, and My Transportation Cost Calculator (MTCC), which takes user-input information on household income, size, and number of workers and uses the LAIM to generate customized transportation cost estimates using the household's tenure, cars, employment locations, and travel patterns.

In addition to updating all the constituent data sources The *Location Affordability Index Model Version 2* (LAIM Version 2) represented a significant a methodological and technical advance from LAIM Version 1. LAIM Version 1 estimated three variables for transportation behavior (auto ownership, auto use, and transit use) and housing costs for homeowners and renters using separate Ordinary Least Squares (OLS) regression models. In LAIM Version 2, however, auto ownership, housing costs, and transit usage for both homeowners and renters are modeled concurrently using simultaneous (or structural) equation modeling (SEM) to capture the interrelationship of these factors.<sup>1</sup> The inputs to the SEM model include these six endogenous variables and 18 exogenous variables. As with Version 1, the new 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.

The current release of the model and data – *Location Affordability Index Model Version 3* (LAIM Version 3 or LAIM3) – provides a data set that now covers the entire US and provides a robust, standardized household housing and transportation cost estimates at the Census tract level for the United States. Like in LAIM Version 2, auto ownership, housing costs, and transit usage for both homeowners and renters are modeled concurrently using simultaneous (or structural) equation modeling (SEM) to capture the interrelationship of these factors. The inputs to the SEM model include these six endogenous variables and 18 exogenous variables. As with previous versions, the new model is used to estimate housing and transportation costs for eight different household profiles, to focus on the impact of the built environment on these costs by holding demographic characteristics constant. Table 1 enumerates all the differences between Versions 2 and 3 of the LAIM. These differences are also referenced throughout the document when relevant.

*Table 1: Changes Between Location Affordability Index Model Version 2 and 3*

<b>Attribute</b>	<b>LAIM3</b>	<b>LAIM2</b>
ACS data vintage	2016 5-Year ACS	2012 5-Year ACS
LEHD data vintage	2014 <sup>2</sup>	2010

<sup>1</sup> Limitations of the data for VMT did not allow for its inclusion in the SEM; it continues to be modeled in Version 2 and Version 3 using OLS.

<sup>2</sup> LEHD data is unavailable for 2014 in Wyoming. So, LEHD data for 2013 is used instead.

Level of geographical granularity	Tract	Block Group
Catchment area for Local Job Density and Local Retail Density (variables 6 and 7)	Used simply the number of workers in the tract and the land area	Used ½ mile buffer around centroid of Block Group and take the union with the BG and used that geography to get Employees and land area
National Transit Database (NTD) vintage	2014	2010
VMT data vintage	2013-2015	2008-2010
Allocation of Fare Box revenue from NTD	Used NTD Primary Urbanized Area	Used AllTransit™ stops and frequency to allocate revenue and trips
Region of Transit Service provider in NTD	Urbanized Area	Metro/Micro Area
When BG/Tract not in Transit service area how to estimate $\alpha$ and $\beta$	Used the values from the nearest urbanized area that had good data	Used the national average
Linearization Functions	<p>Choose the function that gave on average the best OLS fit for all six endogenous variables resulting in different functions for the following variables (see Table 5 below):</p> <ul style="list-style-type: none"> <li>• Owners Average Household Size</li> <li>• Renter Average Household Size</li> <li>• Gross HH Density</li> <li>• Employment Access Index</li> <li>• Fraction of Single Family Detached HU</li> <li>• Fraction of Rental HU</li> <li>• Local Retail Jobs Density</li> </ul>	Optimized each exogenous variable by finding which function made the variables distribution the most normal
Endogenous Variable Interactions	Included interactions if they improved goodness of fit (see Table 3)	Included interactions based primarily on significance
Top and Bottom Code housing costs	No bottom or top coding	Bottom coded to 10th percentile housing cost (owner or renter) within BG and top coded to 90th percentile.

Household Income as Percentile of Tract Income	Included for each household profile	Not included
VMT model	Used odometer readings averaged over Tract	Used odometer readings averaged over Block Group

HUD would like to acknowledge Avar Consulting Inc. and the Center for Neighborhood Technology for their work on and commitment to this project.

## II. LAIM Version 3

### A. Basic Index Structure

Like LAIM Version 2, Version 3 uses an SEM regression analysis to estimate household auto ownership, transit use, and housing costs and a second-order flexible form of ordinary least squares (OLS) model to estimate Vehicle Miles Travelled (VMT). The advantage of using the SEM is that it accounts for the interactions between multiple endogenous variables that are themselves predicted by a set of exogenous variables, which are analogous to the independent variables in an OLS model. The goodness of fit for the SEM measured by a combination of measures in addition to calculated R-squared values for each endogenous variable (see Final Model Structure and Formula below for further discussion on goodness-of-fit measures).

LAIM Version 3 is constructed at the Census tract level using the 2016 American Community Survey (ACS) 5-year estimates as the primary dataset. This is the predominant source for input parameters and measured data for the dependent variables. The LAIM Version 3 covers every occupied Census tract in the 50 states and the District of Columbia.<sup>3</sup>

### B. Data Sources

LAIM Version 3 is produced from data drawn from a combination of the following Federal sources:

- U.S. Census American Community Survey (ACS) – an ongoing survey that generates data on community demographics, income, employment, transportation use, and housing characteristics. 2012-2016 survey data are used in LAIM3.
- U.S. Census TIGER/Line Files – contains data on geographical features 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, which are built on 2014 Census data, are used here.

These data describe relevant characteristics of every census tract in the United States. Census tracts contain between 2,500 and 8,000 residents and vary in size depending on an area's population density. They range from several city blocks to the entirety of some rural counties.

---

<sup>3</sup> There are a few tracts in the United States that do not contain households and were not modeled.

## C. Variables

Starting with a pool of potential independent (exogenous in the SEM) variables representing all the possible influences on housing and transportation costs for which data were available, exogenous variables for the model were chosen according to the strength of their correlation with the endogenous variables and their statistical significance. The choice of variables for LAI Version 2 builds on the theoretical framework developed for LAIM Version 1 with federal stakeholders and the technical review panel. Table 2 lists the final set of variables used in LAIM Versions 2 and 3, with endogenous variables shaded.<sup>4</sup> The following section describes these variables in detail, grouping them by the major characteristics of the built and social environment that they help describe: household density; street connectivity and walkability; employment access and diversity; housing characteristics; household characteristics; housing costs; and household travel behavior.

*Table 2: Overview of LAIM Variables*

Variable	Description	Data Source
1. Gross Density	# of households (HH) / total acres	Census ACS, TIGER/Line files
2. Block Density	# of blocks / total land area	Census TIGER/Line files
3. Employment Access Index	Number of jobs in area block groups / squared distance of block groups	Census LEHD-LODES
4. Retail Employment Access Index	Number of retail jobs in area block groups / squared distance of block groups	Census LEHD-LODES
5. 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
6. Local Job Density	# of jobs / total land area	Census LEHD-LODES
7. Local Retail Density	# of retail jobs / total land area	Census LEHD-LODES
8. Fraction of Rental Units	Number of rental units as a percentage of total housing units	Census ACS
9. Fraction of Single Family Detached Housing Units	Number of single family detached housing units as a percentage of total housing units	Census ACS
10. Median Rooms/Owner HU	Median number of rooms in owner occupied housing units (HU)	Census ACS
11. Median Rooms/Renter HU	Median number of rooms in renter occupied housing units	Census ACS
12. Area Median Income		
13. Owners Median Income Fraction of Area Median Income	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
14. Renters Median Income Fraction of Area Median Income	Median income for renters 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

<sup>4</sup> The table lists the variables in the order they appear they appear below. Variables occur in a different order in the data set itself.

15. Average Household Size: Owners	Calculated from data on Tenure and Total Population in Occupied Housing Units by Tenure	Census ACS
16. Average Household Size: Renters	Calculated from data on Tenure and Total Population in Occupied Housing Units by Tenure	Census ACS
17. Average Commuters per Household Owners	Calculated using the total number of workers 16 years and over who do not work at home	Census ACS
18. Average Commuters per Household Renters	Calculated using the total number of workers 16 years and over who do not work at home	Census ACS
19. Median Selected Monthly Owner Costs	Includes mortgage payments, utilities, fuel, and condominium and mobile home fees where appropriate	Census ACS
20. Median Gross Rent	Includes contract rent as well as utilities and fuel if paid by the renter	Census ACS
21. Autos per Household Owners	Calculated from Aggregate Number of Vehicles Available by Tenure and Occupied Housing Units	Census ACS
22. Autos per Household Renters	Calculated from Aggregate Number of Vehicles Available by Tenure and Occupied Housing Units	Census ACS
23. Percent Transit Journey to Work Owners	Calculated from Means of Transportation to Work by Tenure	Census ACS
24. Percent Transit Journey to Work Renters	Calculated from Means of Transportation to Work by Tenure	Census ACS

The following detailed descriptions of variables used for LAIM Version 3 are organized according to the seven largest factors that influence transportation costs: density; connectivity and walkability; employment access and diversity; housing characteristics; individual household characteristics; housing costs; and household travel behavior. Appendix 1 shows some of the relationships of the endogenous and exogenous variables.

## 1. Household Density

Household density has been found to be one of the largest factors in explaining the variation in all three transportation dependent variables. Various definitions of density have been constructed and tested, and the following two have been utilized in modeling both housing and transportation costs.

### *Variable 1: Gross Density*

Gross Density is calculated as total households (from the ACS) divided by total land acres (calculated using TIGER/Line files).

## 2. Street Connectivity and Walkability

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 clearly have an impact on the pedestrian environment (e.g., crime), the following measure of street connectivity has been found to be an important driver of auto ownership, auto use, and transit use.



### **Variable 2: Block Density**

Census TIGER/Line files are used to calculate average block density (in acres) using the number of blocks within the tract divided by the total block group land area.

### **3. Employment Access and Diversity**

Employment numbers are calculated using OnTheMap Version 7 which provides Longitudinal Employer-Household Dynamics (LEHD) Origin Destination Employment Statistics (LODES) at the Census block level for 2014. These data are currently unavailable in Wyoming.<sup>5</sup>

Measures of employment access and density provide not only an examination of access to work but are good surrogates for proximity to 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.

### **Variable 3: Employment Access Index**

The Employment Access Index is determined using a gravity model which considers both the quantity of and distance to all employment destinations, relative to any given Census tract. Using an inverse-square law, an employment index is calculated by summing the total number of jobs divided by the square of the distance to those jobs. This quantity allows for the examination of both the existence of jobs and the accessibility of these jobs for a given Census tract. Because a gravity model enables consideration of jobs both directly in and adjacent to a given Census tract, the employment access index gives a better measure of job opportunity, and thus a better understanding of job access than a simple employment density measure. This index also serves as a surrogate for access to economic activity.

The Employment Access Index is calculated as:

$$E \equiv \sum_{i=1}^n \frac{P_i}{r_i^2}$$

Where

$E$  = Employment Access for a given Census tract

$n$  = total number of Census tracts

$P_i$  = number of jobs in the  $i^{\text{th}}$  Census tract

$r_i$  = distance (in miles) from the center of the given Census tract to the center of the  $i^{\text{th}}$  Census tract

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

---

<sup>5</sup> Wyoming data was available for 2013 and that was used instead.

#### ***Variable 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).

#### ***Variable 5: Median Commute Distance***

Median commute distance is calculated using LODES data. Median distances are calculated for each Census block using Euclidean (as the crow flies) distances between the origin and destination Census blocks. Block values are then sorted by distance to obtain the median value for the tract of interest.

#### ***Variable 6: Local Job Density***

Local Job Density is a simple measure of the number of jobs from the LODES data divided by the land area (in acres) within the tract.

#### ***Variable 7: Local Retail Density***

The calculation is done using the same measure as the Local Job Density but limited to only jobs in the retail sector as defined by the 2-digit NAICS code from LODES.

### **4. Housing Characteristics**

Characteristics of the housing stock and tenure have been found to influence household travel behavior. Fraction of Rental Units serves as a measure of tenure within a neighborhood. The model incorporates data on housing stock, specifically percent of single-family detached housing units, to further understand the impact of the built environment on transportation decisions. The 2016 ACS 5-year estimates serve as the data source for variables pertaining to housing characteristics.

#### ***Variable 8: Fraction of Rental Units***

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

#### ***Variable 9: Fraction of Single Family Detached Housing Units***

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

#### ***Variable 10: Median Number of Rooms in Owner-Occupied Housing Units***

Data on Median Number of Rooms by Tenure is determined from the ACS and is included as an exogenous variable. In cases where the Median Number of Rooms in owner occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

#### ***Variable 11: Median Number of Rooms in Renter-Occupied Housing Units***

Data on Median Number of Rooms by Tenure is determined from the ACS and is included as an exogenous variable. In cases where the Median Number of Rooms in renter occupied households is suppressed the value for the tract is used in running the model but not for calibrating the model.

## 5. Household Characteristics

The 2016 ACS 5-year estimates serve as the primary data source for variables pertaining to household characteristics.

### ***Variable 12: Area Median Income***

Median household income is obtained directly from the ACS at the CBSA level for tracts in metropolitan and micropolitan area and at the county level for all other tracts.

### ***Variable 13: Owners Median Income Fraction of Area Median Income***

Fraction of area median income for owners is calculated as the ratio of median income for owners at the tract level to the Area Median Income. In cases where the tract median income for owner occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

### ***Variable 14: Renters Median Income Fraction of Area Median Income***

Fraction of area median income for renters is calculated as the ratio of median income for renters at the tract level to the Area Median Income. In cases where the tract median income for renter occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

### ***Variable 15: Average Household Size Owners***

Average household size for owners is calculated using Tenure and Total Population in Occupied Housing Units by Tenure to define the universe of Owner-occupied Housing Units. The total population in owner units is divided by the number of owner units. In cases where the tract population in owner occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

### ***Variable 16: Average Household Size Renters***

Average household size for renters is calculated using Tenure and Total Population in Occupied Housing Units by Tenure to define the universe of Renter Occupied Housing Units (see paragraph E. iv). In cases where the tract population in renter occupied households is suppressed the value for the tract is used in running the model but not for calibrating the model.

### ***Variable 17: Average Commuters per Household Owners***

Average commuters per household is 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 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 Housing Units to Total Population is used to scale the count of commuters to better represent those living in households. In cases where the tract population in owner occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

### ***Variable 18: Average Commuters per Household Renters***

Average commuters per household is 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 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 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). In cases where the tract population in renter occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

## 6. Housing Costs

The 2012-2016 ACS 5-year estimates serve as the data source for variables pertaining to housing costs.

### *Variable 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.

### *Variable 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, this measure is used to measure Renters Housing Cost.

## 7. Household Transportation Behavior

The 2012-2016 ACS 5-year estimates serve as the data source for variables pertaining to household travel behavior.

### *Variable 21: Autos per Household Owners*

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

### *Variable 22: Autos per Household Renters*

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

### *Variable 23: Percent Transit Journey to Work Owners*

As no direct measure of transit use is available at the tract level, a proxy is utilized for the measured data to represent the variable of transit use. From the ACS, Means of Transportation to Work by Tenure is used to calculate a percent of commuters in owner-occupied housing utilizing public transit.

### *Variable 24: Percent Transit Journey to Work Renters*

As no direct measure of transit use is available at the tract level, a proxy is utilized for the measured data to represent the variable of transit use. From the ACS, Means of Transportation to Work by Tenure is used to calculate a percent of commuters in renter-occupied housing utilizing public transit.

## D. Simultaneous Equations Model

The following description applies to both LAIM2 and LAIM3. The table below summarizes methodological differences between the two versions of the model; see Appendix C for a review of the development of LAIM2.

## 1. 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:

- Owner Auto Ownership
- Renter Auto Ownership
- Renters Housing Cost – measured by Gross Rent
- Owners Housing Cost – measured by Selected Monthly Ownership Costs (SMOC)
- Owner Transit Commute Share
- Renter Transit Commute Share

Figure 1 (next page) shows a schematic representation of the relationships in the SEM between endogenous variables for LAIM3 (note that these are different from LAIM2). Causality can go both ways; often it was found that once causality is explained in one direction, the other direction is either not statistically significant or markedly less significant and in LAIM2 we did not include these; however, in LAIM3 we found that some of these interactions provided a better goodness of fit and were included.

Table 3 (following page) shows the relationships in the final model and their hypothesized mechanisms. 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 Renter Transit Commute Share or Owners Housing Cost).

Figure 1: Schematic Representation of the Relationships between the Endogenous Variable Implemented in the SEM

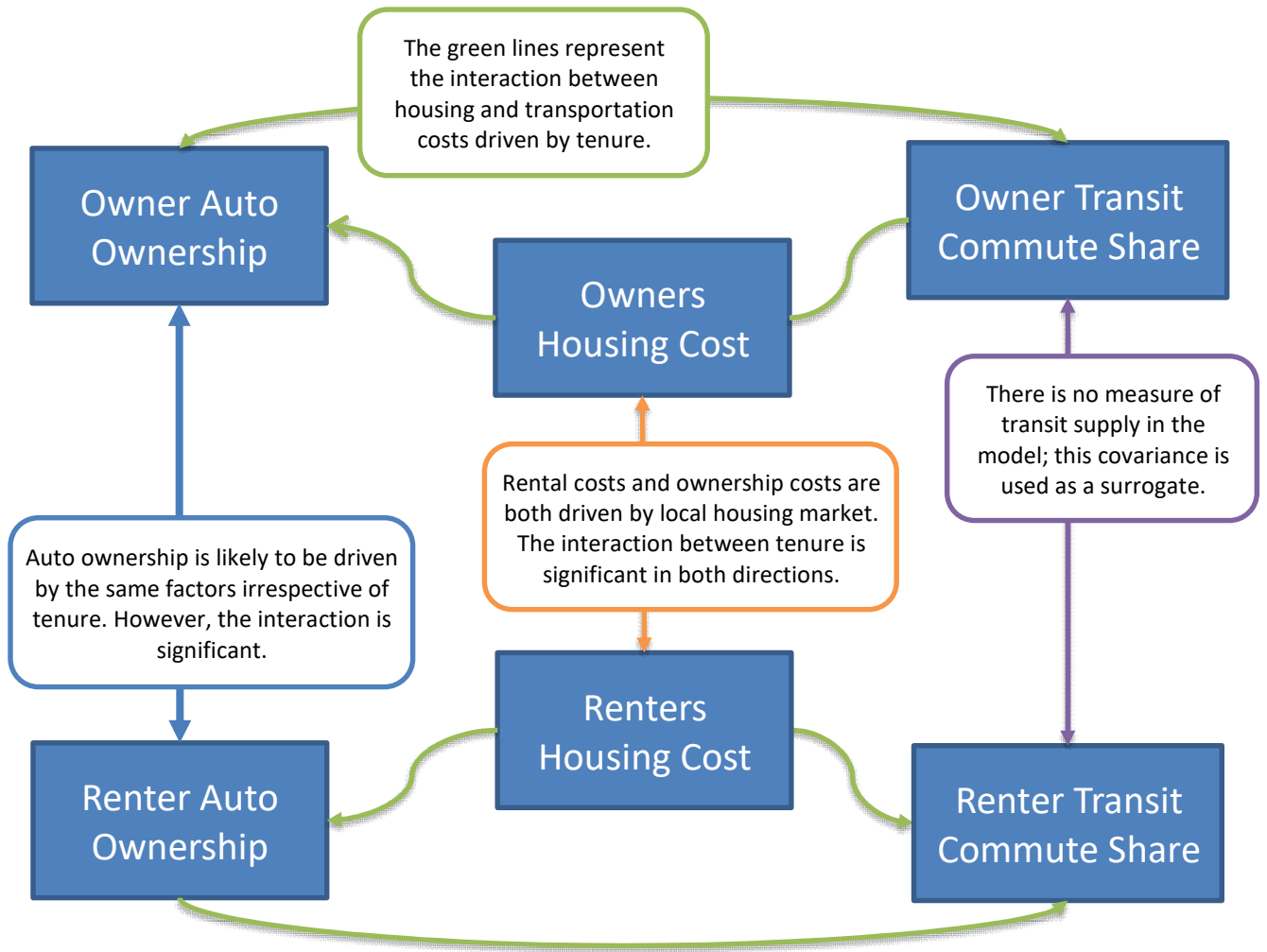


Table 3: Endogenous Variable Interactions – Hypothesized Mechanisms

Variable 1 (V <sub>1</sub> )	Variable 2 (V <sub>2</sub> )	Hypothesized Mechanism	Interaction Used in LAIM3	Interaction Used in LAIM2
Owner Auto Ownership	Renter Auto Ownership	Auto ownership is driven by many of the same factors independent of tenure. The correlation observed here is presumably more coincidental than causal; however, linking the two enhances the predictions for both capturing this residual effect.	Two Way (V <sub>1</sub> ↔ V <sub>2</sub> )	None
Owner Auto Ownership	Owners Housing Cost	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 <sub>2</sub> → V <sub>1</sub> )	One Way (V <sub>2</sub> → V <sub>1</sub> )
Owner Auto Ownership	Owner Transit Commute Share	Auto ownership and transit use are obviously related.	Two Way (V <sub>1</sub> ↔ V <sub>2</sub> )	One Way (V <sub>1</sub> → V <sub>2</sub> )
Renter Auto Ownership	Renters Housing Cost	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 <sub>2</sub> → V <sub>1</sub> )	One Way (V <sub>2</sub> → V <sub>1</sub> )
Renter Auto Ownership	Renter Transit Commute Share	Auto ownership and transit use are obviously related.	One Way (V <sub>1</sub> → V <sub>2</sub> )	One Way (V <sub>1</sub> → V <sub>2</sub> )
Owners Housing Cost	Renters Housing Cost	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.	Two Way (V <sub>1</sub> ↔ V <sub>2</sub> )	One Way (V <sub>1</sub> → V <sub>2</sub> )
Owners Housing Cost	Owner Transit Commute Share	The cost of transit is relatively low (compared to auto ownership) thus the constraint driven by a household's budget is less rigid. No strong reason interaction was observed.	None	None
Renters Housing Cost	Renter Transit Commute Share	The cost of transit is relatively low (compared to auto ownership), but renters' budgets (i.e. incomes) are lower and thus transit costs have a greater impact than for owners.	One Way (V <sub>1</sub> → V <sub>2</sub> )	None

Owner Transit Commute Share	Renter Transit Commute Share	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 ( $V_1 \leftrightarrow V_2$ )	Two Way ( $V_1 \leftrightarrow V_2$ )
-----------------------------	------------------------------	--	---------------------------------------	---------------------------------------

## 2. Variable Transformation

For LAIM2, we tested a variety of transformation to see which one had the normal distribution for every variable; this method presumes that the underlying distribution is normal. However, this technique will not necessarily account for non-linearity in the relationships between the dependent (endogenous) and independent (exogenous) variables, giving the highest  $R^2$  in an OLS. So, to optimize the transformation for LAIM3, for each exogenous variable we used six OLS models for each of the endogenous variables using the appropriate (by tenure) exogenous variable and cycling through each variable using one of six transformation formulae. Since this model uses many exogenous variables to predict six endogenous values, the optimal transformation may be different for each, that is why the transformation that gives the best average  $R^2$  is then assigned to each exogenous variable. Table 4 shows an example of this for the Employment Access Index showing that the previous functional form of natural log ( $\ln(x)$ ) was on average not the optimal transformation, and by changing it to the square root ( $\sqrt{x}$ ) would on average improve the fits.

Table 4:  $R^2$  for OLS fit using all variables, but changing the transformation function for  $x$ =Employment Access Index

Endogenous Variable Function	Owner Auto Ownership	Renter Auto Ownership	Renter Housing Cost	Owner Housing Cost	Owner Transit Commute Share	Renter Transit Commute Share	Average
x - Linear	69.39%	67.90%	70.72%	77.51%	59.82%	60.43%	<b>67.63%</b>
sqrt(x) – Square Root	69.39%	67.84%	70.73%	77.56%	60.89%	62.29%	<b>68.12%</b>
ln(x) – Natural Log	69.47%	67.63%	70.68%	77.74%	57.05%	58.53%	<b>66.85%</b>
ln(1+x) – Safe Natural Log	69.47%	67.63%	70.68%	77.74%	57.05%	58.53%	<b>66.85%</b>
1/x – Inverse	69.21%	67.46%	70.70%	77.83%	57.50%	58.72%	<b>66.90%</b>
1/(1+x) – Safe Inverse	69.21%	67.46%	70.70%	77.83%	57.51%	58.73%	<b>66.91%</b>



After proceeding through all the endogenous variables a few of linearization formula were changed to a new optimal form. Table 5 show these changes that are implemented in a new LAIM3 versus LAIM2:

*Table 5: Linearization transformation functions from LAIM3 and LAIM2*

Variable	LAIM3	LAIM2
Owners Fraction of AMI	ln(x)	
Renter Fraction of AMI	ln(x)	
Area Median Income (AMI)	ln(x)	
Owners Average Household Size	x	ln(x)
Renter Average Household Size	x	ln(x)
Block Density	$\sqrt{x}$	
Owner Commuters/HH	x	
Renter Commuters/HH	x	
Gross HH Density	x	$\sqrt{x}$
Local Job Density	$\sqrt{x}$	
Employment Access Index	$\sqrt{x}$	ln(x)
Median Commute Distance	ln(x)	
Owners Rooms/HU	x	
Renter Rooms/HU	x	
Fraction of Single Family Detached HU	$\sqrt{x}$	x
Fraction of Rental HU	x	$\sqrt{x}$
Local Retail Jobs Density	ln(1+x)	$\sqrt{x}$
Retail Employment Access Index	ln(x)	
Owner Auto Ownership	x	
Renter Auto Ownership	x	
Renter Housing Cost	ln(x)	
Owner Housing Cost	ln(x)	
Owner Transit Commute Share	x	
Renter Transit Commute Share	x	

### 3. Variable Standardization

Finally, all variables going into the SEM are transformed using the familiar *standard score* such that they can be measured on the same scale and used in the same functions.<sup>6</sup>

Table 6: Variables Used to Estimate the Model, with Transformations and Descriptive Statistics

Variable Name	Transformation	Transformed Mean	Transformed Standard Deviation
Owners Auto Ownership	Linear	1.981	0.356
Renters Auto Ownership	Linear	1.372	0.393
Owners Transit Commute Share	Linear	3.993	9.902
Renters Transit Commute Share	Linear	5.968	12.717
Renters Housing Cost	Natural Log	6.816	0.340
Owners Housing Cost	Natural Log	7.233	0.360
Renters Fraction of AMI	Natural Log	-0.454	0.398
Renters Commuters/HH	Linear	1.099	0.332
Median Commute Distance	Natural Log	2.567	0.576
Owners Rooms/HU	Linear	6.221	0.859
Owners Fraction of AMI	Natural Log	0.163	0.316
Area Median Income (AMI)	Natural Log	10.910	0.211
Owners Household Size	Linear	2.692	0.520
Renters Household Size	Linear	2.597	0.665
Block Density	Square Root	0.265	0.161
Owners Commuters/HH	Linear	1.202	0.289
Gross HH Density	Linear	2.894	6.404
Employment Access Index	Square Root	142.620	95.141
Renters Rooms/HU	Linear	4.577	0.803
Fraction of Single-Family Detached HU	Square Root	7.601	2.023
Fraction of Rental HU	Linear	33.044	18.766
Retail Employment Access Index	Natural Log	7.311	1.267
Local Job Density	Square Root	1.119	1.202
Local Retail Job Density	Safe Natural Log	0.182	0.302

HU = Housing Units                      HH = Households  
 Endogenous variables are shaded.

<sup>6</sup> The standardized score (Z) for each data point of every variable is equal to the value of that data point minus the mean for that variable, divided by the variable's standard deviation. Z is thus equal to the number of standard deviations by which the value of the data point differs from the mean for each variable.

This standardization technique—converting to z-scores—was applied to each variable to enable the SEM function in R<sup>7</sup> to handle the wide variation in values. 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.

#### 4. Final Model Structure and Formula

As previously mentioned, the SEM used in LAIM Version 3 consists of six nested equations, each drawing from a pool of 18 exogenous variables, that predict six interrelated endogenous variables. The variables chosen for inclusion in LAIM3 are the same as LAIM2 – this was a decision at the outset of LAIM3 development to keep the LAI consistent.

Table 6 shows the structure of the SEM model used in LAIM3, organized by the six nested equations, one corresponding to each of the model’s six endogenous variables (numbered and including the final R<sup>2</sup>). All endogenous variables appearing as exogenous variables in other nested equations are shaded as well.

Table 6: SEM Structure<sup>8</sup>

Variables	Estimate	Std. Error	t-Value
<b>1. Owners Auto Ownership (R<sup>2</sup> = 72.38%)</b>			
Fraction of Single Family Detached HU	0.262	0.004	70.1
Owners Commuters/HH	0.258	0.004	69.1
Block Density	-0.190	0.003	55.0
Fraction of Rental HU	0.175	0.003	51.6
Renters Auto Ownership	0.150	0.003	44.0
Gross HH Density	-0.127	0.003	40.9
Owners Fraction of AMI	0.151	0.004	39.7
Retail Employment Access Index	-0.190	0.005	38.3
Owners Household Size	0.128	0.003	37.1
Area Median Income (AMI)	0.089	0.004	23.2
Employment Access Index	-0.119	0.005	22.0
Owners Rooms/HU	0.061	0.003	21.5
Median Commute Distance	0.058	0.003	20.2
Owners Housing Cost	0.048	0.004	11.0
<b>2. Renters Auto Ownership (R<sup>2</sup> = 71.27%)</b>			
Renters Commuters/HH	0.326	0.003	100.7
Renters Fraction of AMI	0.177	0.003	58.6
Owners Auto Ownership	0.188	0.004	48.9
Renters Housing Cost	0.144	0.003	43.6
Renters Rooms/HU	0.115	0.003	39.6

<sup>7</sup> R is a software programming language used for statistical analysis.

<sup>8</sup> Nested equations with R<sup>2</sup> values are given for each of the numbered exogenous variables with estimated value, standard error, and t-value for each included endogenous variable. Exogenous variables appearing as endogenous variables are shaded.

<b>Variables</b>	<b>Estimate</b>	<b>Std. Error</b>	<b> t-Value </b>
Fraction of Single Family Detached HU	0.103	0.003	29.7
Employment Access Index	-0.153	0.006	27.7
Gross HH Density	-0.086	0.003	27.1
Block Density	-0.097	0.004	26.8
Renters Household Size	0.064	0.003	21.8
Local Job Density	0.055	0.003	16.8
Retail Employment Access Index	-0.081	0.005	16.3
<b>3. Renters Housing Cost (R<sup>2</sup> = 75.64%)</b>			
Renters Fraction of AMI	0.375	0.003	136.8
Area Median Income (AMI)	0.326	0.003	104
Owners Housing Cost	0.309	0.003	88.6
Retail Employment Access Index	0.321	0.004	82.0
Renters Household Size	0.153	0.003	57.2
Renters Rooms/HU	0.107	0.002	43.5
Employment Access Index	-0.094	0.004	23.8
Renters Commuters/HH	-0.055	0.003	18.1
<b>4. Owners Housing Cost (R<sup>2</sup> = 78.48%)</b>			
Owners Fraction of AMI	0.561	0.002	241.8
Area Median Income (AMI)	0.539	0.003	182.9
Owners Household Size	0.216	0.003	74.5
Owners Commuters/HH	-0.208	0.003	66.1
Employment Access Index	0.189	0.003	58.9
Fraction of Single Family Detached HU	-0.108	0.003	36.3
Renters Housing Cost	0.092	0.003	27.5
Median Commute Distance	0.064	0.002	27.1
Fraction of Rental HU	-0.061	0.003	21.1
<b>5. Owners Transit Commute Share (R<sup>2</sup> = 73.37%)</b>			
Owners Auto Ownership	-0.305	0.004	74.7
Gross HH Density	0.255	0.004	70.3
Employment Access Index	0.356	0.007	50.8
Retail Employment Access Index	-0.234	0.005	44.6
Local Job Density	-0.132	0.004	34.6
Owners Household Size	0.109	0.003	34.1
Fraction of Rental HU	-0.107	0.003	32.5
Renters Transit Commute Share	0.246	0.008	32.4
Owners Housing Cost	0.085	0.003	26.7
Fraction of Single Family Detached HU	-0.083	0.004	21.7
Area Median Income (AMI)	0.067	0.003	21.2
Local Retail Jobs Density	0.057	0.003	19.5
Owners Commuters/HH	0.067	0.004	19.3
Median Commute Distance	0.049	0.003	17.4

Variables	Estimate	Std. Error	t-Value
Owners Rooms/HU	0.044	0.003	17.1
<b>6. Renters Transit Commute Share (R<sup>2</sup> = 77.85%)</b>			
Renters Auto Ownership	-0.274	0.003	82.4
Owners Transit Commute Share	0.434	0.006	69.3
Employment Access Index	0.306	0.006	51.5
Local Job Density	-0.096	0.003	32.4
Retail Employment Access Index	-0.146	0.005	32.4
Area Median Income (AMI)	0.086	0.003	32.0
Renters Household Size	0.075	0.002	30.3
Fraction of Single Family Detached HU	-0.086	0.003	29.6
Renters Rooms/HU	0.073	0.003	28.7
Gross HH Density	0.101	0.004	28.5
Median Commute Distance	-0.045	0.003	18.0
Renters Commuters/HH	0.051	0.003	17.7
Renters Housing Cost	-0.009	0.003	3.1

See Appendix 2: for a path diagram that visualizes the relative strength of these correlations. Table 7 enumerates the nature and strength of the salient relationships between the model's endogenous variables.

Table 7: Relationships of the Endogenous Variables

Endogenous Variable 1	Endogenous Variable 2	Value of Coefficient (for transformed and normalized variables)	Nature of relationship
Owners Auto Ownership	Renters Auto Ownership	0.150 +/- 0.003	As renter own more autos, so to home owners.
Owners Auto Ownership	Owners Housing Cost	0.048 +/- 0.004	As home ownership costs go up, auto ownership increases.
Renters Auto Ownership	Owners Auto Ownership	0.188 +/- 0.004	As home owners own more autos, so do renters.
Renters Auto Ownership	Renters Housing Cost	0.143 +/- 0.003	As rents goes up, auto ownership increases for renters.
Renters Housing Cost	Owners Housing Cost	0.309 +/- 0.003	As home ownership costs go up, rents increase.
Owners Housing Cost	Renters Housing Cost	0.092 +/- 0.003	As rents go up, home ownership costs increase, but not as fast as the other way around.
Owner Transit Commute Share	Owners Auto Ownership	-0.304 +/- 0.004	As auto ownership goes up, transit ridership decreases for home owners.
Owner Transit Commute Share	Owners Housing Cost	0.085 +/- 0.003	As housing costs goes up, transit ridership increases for home owners.
Owner Transit Commute Share	Renter Transit Commute Share	0.247 +/- 0.008	As more owners use transit, more renters do as well.

Renter Transit Commute Share	Renters Auto Ownership	-0.274 +/- 0.003	As auto ownership goes up, transit ridership decreases for renters.
Renter Transit Commute Share	Renters Housing Cost	-0.009 +/- 0.003	As housing costs go up, transit ridership decreases for renters (but only slightly).
Renter Transit Commute Share	Owner Transit Commute Share	0.433 +/- 0.006	As more renters use transit, more owners do as well.

The complexity of SEMs has resulted in a range of metrics to assess the model goodness of fit. For the SEM employed in LAIM3, 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):** RMSEA 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. The rule of thumb that Kline reports is that an “RMSEA  $\leq$  0.05 indicates close approximate fit, values between 0.05 and 0.08 suggest reasonable error of approximation, and RMSEA  $\geq$  0.10 suggests poor fit.” A 90% confidence interval is commonly used to assess the range of the RMSEA score. The LAIM3 model has an RMSEA of 0.062 whose 90% confidence interval ranges from 0.061 to 0.063.
2. **Comparative Fit Index (CFI):** CFI 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. The rule of thumb that Kline reports is that CFI “values greater than roughly 0.90 may indicate reasonably good fit of the researcher’s model.” The LAIM3 model has a CFI of 0.973.
3. **Standardized Root Mean Square Residual (SRMR):** SRMR 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. The rule of thumb Kline reports is that “values of the SRMR less than 0.10 are generally considered favorable.” The LAIM3 model has an SRMR of 0.013.

By achieving these three robust measures, the SEM model used for LAIM Version 3 is shown to be an efficient statistical model.

The three metrics above relate to the overall quality of the SEM. The R<sup>2</sup> values for each endogenous variable used in the model are listed in Table 8.

*Table 8: Summary of R<sup>2</sup> Values for each Endogenous Variable*

Endogenous Variable	R <sup>2</sup>
Owners Auto Ownership	72.38%
Renters Auto Ownership	71.27%
Renters Housing Cost	75.64%
Owners Housing Cost	78.48%
Owners Transit Commute Share	73.37%
Renters Transit Commute Share	77.85%

## E. Modeling Vehicle Miles Traveled

As noted previously, auto use—measured in Vehicle Miles Travelled or VMT—cannot be included in the SEM because VMT data is only available for the state of Illinois; VMT is instead modeled using OLS regression. The regression model was fit using data on the total number of miles households that drive their autos, calculated from odometer readings from the Chicago and St. Louis metro areas for 2013 through 2015, obtained from the Illinois Environmental Protection Agency. Two odometer readings—for 2013 and 2015—were matched for over 900,000 vehicles using vehicle identification numbers (VIN) to obtain data for VMT during that period.

Although limited to Illinois, the geographic area that the data covers includes a variety of place types—from rural to large city—which provides excellent fodder for calibrating a model. To assess the validity of this data set for use predicting VMT for the entire country, national driving records were obtained from the National Household Travel Survey (NHTS) and assigning them to Census tracts using ZIP+4<sup>TM</sup> geographical identifications. Automobiles were matched using their VIN and the total distance driven was determined over the time between inspections. The resulting analysis showed that the ratio of the average ANNMILES to the average VMT predicted by the LAI VMT model by census region<sup>9</sup> was 1.08, suggesting that the LAI VMT model slightly underestimates auto usage nationwide. Previous analysis suggests that most of this discrepancy is due to the vehicles included in the Illinois EPA data all being at least five years old, and in the aggregate older cars are driven less than newer ones. To compensate, the final values of VMT given in the model include an adjustment factor of eight percent.

In both versions of the LAIM, VMT is predicted using OLS regression analysis with a second-order flexible functional form. This functional form allows us to take into consideration the interactions between independent variables in addition to their unique effects; for instance, household density, household income, and the combination of the two 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. The difference is that this VMT model is run once for each household profile irrespective of tenure, so overall average income, household size and commuters per household were used rather than two tenure-specific versions of each variable.

Additionally, because there is an inherent spatial autocorrelation for the dependent variables, it is necessary to employ a robust spatial variance calculation in estimating the resulting error in each regression coefficient, allowing better evaluation of their statistical significance and helping determine whether each specific variable or variable combination was used in the final fit. To determine the best estimator of the error on model coefficients due to spatial autocorrelation, we tested geographical clustering at the state, county, and CBSA levels. Results showed that as expected the estimated errors increased when using this approach, with county-level and CBSA-level clustering having similar effects on error and state-level clustering having a lesser effect than either. Consequently, CBSA clustering was used for CBSAs and county-level clustering was used for non-CBSA areas.

There is a high probability that the independent variables are multi-collinear, so after eliminating coefficients with high p-value, the coefficients with Variance Inflation Factor (VIF) of 100 or greater were also dropped.<sup>10</sup> This was done incrementally, once the variables that are not statistically significant are dropped the VIF is examined; the VIF values for this analysis tended to be greater than 10,000 for the

---

<sup>9</sup> Data were averaged across each Census region (i.e. Midwest, Northeast, South, and West) due to the relatively small sample size of the NHTS.

<sup>10</sup> For a definition of VIF see [http://en.wikipedia.org/wiki/Variance\\_inflation\\_factor](http://en.wikipedia.org/wiki/Variance_inflation_factor) .

most multicollinear variables, but once those are removed from the fit the values drop perceptibly as highly multi-collinear coefficients were excluded, until they are less than 100.

Table 9 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 very collinear with the other variables.

The entire set of cross terms used in the models with their coefficients and values can be found in Table 10: Regression Coefficients for VMT Model on the next page. Note that there is no significant relationship with local job density. This result leads to a need of only one model run per household type since there is no dependence on tenure.

*Table 9: Independent Variables Used in VMT Regression*

Variable	Linear Transformation	Number of Times Used in Combination
Fraction of AMI	Natural Log	3
Area Median Income	Natural Log	2
Household Size	Linear	2
Block Density	Square Root	3
Commuters/HH	Linear	8
Gross HH Density	Linear	3
Local Job Density	Square Root	4
Employment Access Index	Square Root	2
Median Commute Distance	Natural Log	4
Rooms/HU	Linear	2
Fraction of Single Family Detached HU	Square Root	3
Fraction of Rental HU	Linear	2
Local Retail Jobs Density	Safe Natural Log	0
Retail Employment Access Index	Natural Log	2

*Table 10: Regression Coefficients for VMT Model in order of R<sup>2</sup> reduction*

Variable Combination	Value	Standard Error	VIF
Intercept	12373	1767	0.0
Commuters/HH   Fraction of Single Family Detached HU	-584	104	91.8
Block Density   Rooms/HU	-1811	172	16.5
Median Commute Distance	-4756	825	75.1



Commuters/HH   Rooms/HU	555	81	20.3
Retail Employment Access Index   Retail Employment Access Index	-178	21	55.2
Fraction of Rental HU   Fraction of Rental HU	0.3	0.1	15.3
Median Commute Distance   Retail Employment Access Index	810	110	46.7
Household Size   Fraction of Single Family Detached HU	117	26	24.7
Fraction of AMI   Fraction of AMI	-1242	169	2.5
Block Density   Block Density	4127	986	18.6
Commuters/HH   Gross HH Density (e   f)	-52	19	28.5
Block Density   Gross HH Density	77	15	23.8
Gross HH Density   Median Commute Distance	-28	13	61.8
Commuters/HH   Fraction of Rental HU	-40	7	17.4
Area Median Income   Commuters/HH	964	110	63.2
Area Median Income   Fraction of Single Family Detached HU	76	12	59.1
Fraction of AMI   Local Job Density	155	86	4.9
Commuters/HH   Employment Access Index	-12	2	28.9
Employment Access Index   Employment Access Index	0.016	0.002	12.9
Local Job Density   Median Commute Distance	-278	63	21.4
Household Size   Local Job Density	440	84	43.4
Commuters/HH   Local Job Density	-479	135	36.8
Fraction of AMI   Commuters/HH	455	198	6.8

### III. Using the LAIM to Generate the Location Affordability Index (LAI)

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 “selected household”) in the Model’s outputs to control for any variation they might cause. By establishing and running the model for a “selected household,” any variation observed in housing and transportation costs may be attributed to place and location, rather than household characteristics.

#### A. Modeling Transportation Behaviors and Housing Costs

The model was run for the eight household profiles 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 Table 11. They are not intended to match the characteristics of any 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 profiles are based on the median household income for each Combined Base Statistical Area (CBSA) covered by the index, or in the case of non-metropolitan counties, the median household income for the county, making the results regionally specific. It was run for both owner and renter tenure for each type.

Table 11: LAI Household Profiles

Household Profile	Income	Size	Number of Commuters
1. Median-Income Family	MHHI	4	2
2. Very Low-Income Individual	National poverty line	1	1
3. Working Individual	50% of MHHI	1	1
4. Single Professional	135% of MHHI	1	1
5. Retired Couple	80% of MHHI	2	0
6. Single-Parent Family	50% of MHHI	3	1
7. Moderate-Income Family	80% of MHHI	3	1
8. Dual-Professional Family	150% of MHHI	4	2

MHHI = Median household income for a given area (CBSA or County).

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

1. It was applied to both owners and renters. This was done by using the database values for each tract for all the variables that apply to the other tenure (i.e., renters when running owner household, and owners when running renter households – see Table 12).
2. The VMT model was run for each household type, irrespective of tenure.
3. Calculate the transportation cost, for each household type and tenure, using the cost developed for LAI Version 1, but multiply by an inflation factor to determine 2016 dollars from the 2010 calculations.
4. Put costs together with the ratio of each household type income and integrate into the database.

Table 12: Household Variables used in SEM

Modeled Variables	Owner Household Variables <sup>11</sup>	Renter Household Variables <sup>12</sup>
<ul style="list-style-type: none"> <li>• Owners Auto Ownership</li> <li>• Owners Housing Cost</li> <li>• Owner Transit Commute Share</li> </ul>	Values from Table 9	Values from renter households in tract
<ul style="list-style-type: none"> <li>• Renters Auto Ownership</li> <li>• Renters Housing Cost</li> <li>• Renter Transit Commute Share</li> </ul>	Values from owner households in tract	Values from Table 9

## B. Using Modelled Transportation Behavior to Calculate Transportation Costs

As discussed, LAIM Version 3 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 tract. This operation is performed for the estimates generated for each of the eight household types.

<sup>11</sup> Household Income Owners, Household Size Owners, and Commuters per Household Owners

<sup>12</sup> Household Income Renters, Household Size Renters, and Commuters per Household Renters

## 1. 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 Version 3. Research conducted by Diane Schanzenbach, PhD and Leslie McGranahan PhD<sup>13</sup>, which included a range of new and used autos, examined expenditures based on the 2005-2010 waves of the CES. This research advanced the effort to overcome 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). Expenses are segmented by five ranges of household income (<\$20,000; \$20,000-\$39,999; \$40,000-\$59,999; \$60,000-\$99,999; and, \$100,000 and above) and applied to the modeled autos per household and annual VMT for the appropriate income range. LAI Version 3 uses an additional inflation factor of 1.05765<sup>14</sup> to adjust to 2016 dollars.

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

- Average annual service flow value<sup>15</sup> 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.

*Table 13: 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 (<\$20,000)	\$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 calculation of auto cost is:

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

Where

A = Modeled autos per household

V<sub>sf</sub> = Per vehicle service flow cost from Table 13 (1) – for the appropriate income group

V<sub>fc</sub> = Per vehicle finance charge from Table 13 (2) – for the appropriate income group

V<sub>fixed</sub> = Per vehicle (fixed) ownership cost from Table 13 (3) – for the appropriate income group

VMT = the modeled annual household VMT

MPG = the national average fuel efficiency (21.6 mpg for 2014)

G = the cost of gas per gallon (average annual regional cost for 2014)<sup>16</sup>

<sup>13</sup> <https://www.hudexchange.info/resources/documents/LAI-Auto-Cost-Research-Synthesized.pdf>

<sup>14</sup> [http://www.bls.gov/data/inflation\\_calculator.htm](http://www.bls.gov/data/inflation_calculator.htm)

<sup>15</sup> Service flow is the average annual dollar amount of depreciation the vehicle has lost over the time of ownership.

<sup>16</sup> U.S. Department of Energy, Energy Information Administration. "Petroleum & Other Liquids." Accessed from <http://www.eia.gov/petroleum/gasdiesel/>.

R = the Average Ratio drivability to fuel cost from Table 13 (7) – for the appropriate income group

## 2. Transit Use Costs

Transit cost data—directly operated and purchased transportation revenue as reported by each transit agency—were obtained from the 2014 National Transit Database (the middle year of the 2016 5-year ACS dataset).<sup>17</sup> Specifically, we looked at in the database. By employing certain assumptions to allocate farebox revenue to urbanized areas from the National Transit Database (NTD), we can sub-allocate all costs to households within Census tracts contained in those urbanized areas such that the household expenditure for transit equals the agencies revenue.

The NTD assigns a primary urbanized area (UZA) to every transit agency, as well as other non-primary urbanized areas and non-urbanized areas where applicable (for example “Texas Non-UZA”). Table 14 shows that of the 849 transit agencies submitting data to the NTD, most (657) provide service in only one urbanized area, with a subset (280) of those serving non-urbanized areas as well. 191 transit agencies serve multiple urbanized areas, with the majority (155) of those most also serving non-urbanized areas. Only one agency serves only non-urbanized areas.

*Table 14: NTD Transit Agencies by Number of Service Areas*

Number of Service Areas	Transit Agencies
One UZA	377
One UZA and adjacent non-urbanized areas	280
Multiple UZAs	36
Multiple UZAs and adjacent non-urbanized areas	155
Non-urbanized areas only	1
<b>TOTAL</b>	<b>849</b>

Given that there is no ubiquitous and standardized data on station location and service frequency, the allocation of transit agency revenue and trips must be done at the urbanized area geography. Non-urbanized areas are not included in the full revenue allocation since it is impossible to divide the revenue/trips across multiple areas served. Also, this method assumes that the non-primary urbanized areas will average out. Since most transit is provided and used in urban settings there should be a minimal amount of error associated with this allocation method.

We calculate two normalization factors ( $\alpha$  and  $\beta$ ) that estimate the average household’s transit cost ( $\alpha$ ) and trips on transit ( $\beta$ ) given the fraction of commuters in a given Census tract using transit for their

<sup>17</sup> <https://www.transit.dot.gov/ntd/ntd-data>. Demand response revenue is not factored into this analysis.

journey to work.<sup>18</sup> This is accomplished using the total revenue from the transit agencies that serve a given UZA, and the total number of journey-to-work trips in that UZA.<sup>19</sup>

### C. The Location Affordability Index database

The database generated by the foregoing methodology contains records for 72,241 U.S. Census tracts in the 50 states and the District of Columbia. The first set of fields (columns A-AL) contain data used to calibrate the model and then serve as inputs into the model to estimate household housing and transportation costs for the eight household profiles—including estimates for both owner and renter households—listed in Table 11 for each occupied Census tract.

Table 15: LAI Data Dictionary: Model Inputs

Column Name	Description
stfid	Census id
households	Number of households
owner_occupied_hu	Number of owner-occupied housing units
renter_occupied_hu	Number of renter-occupied housing units
pct_renters	Percent of rental housing units
pct_renter_occupied_hu	Percent of households in renter-occupied housing units
pct_transit_j2w_renters	Percent of commuters living in rental households using transit for their journey to work
pct_transit_j2w_owners	Percent of commuters living in owner households using transit for their journey to work
pct_transit_j2w	Percent of commuters using transit for their journey to work
median_smoc_mortgage	Median selected monthly ownership costs
median_gross_rent	Median gross rent
avg_h_cost	Average monthly housing cost
autos_per_hh_renters	Autos per household for renter households
autos_per_hh_owner	Autos per household for owner households
autos_per_hh	Autos per household
commuters_per_hh_renters	Average number of commuters per household in renter households
commuters_per_hh_owners	Average number of commuters per household in owner households
commuters_per_hh	Average number of commuters per household
avg_hh_size_renters	Average number of people in renter households
avg_hh_size_owners	Average number of people in owner households
avg_hh_size	Average number of people per household
area_income_renter_frac	Fraction of area median household income for renters in this tract relative to the regional median household income
median_hh_income	Median household income
median_rooms_per_renter_hu	Median number of rooms in renter households
median_rooms_per_owner_hu	Median number of rooms in owner households
median_rooms_per_hu	Median number of rooms per household

<sup>18</sup> This assumes that fraction of transit used in the journey to work is a good surrogate for all transit use.

<sup>19</sup> For a complete derivation of how  $\alpha$  and  $\beta$  are calculated see the more detailed method paper.

pct_hu_1_detached	Percent of single family detached housing units
gross_hh_density	Number of households per land acre
area_income_owner_frac	Fraction of area median household income for owners in this tract relative to the regional median household income
area_income_frac	Fraction of area median household income in this tract relative to the regional median household income
area_median_hh_income	Regional median household income
block_density	Census blocks per acre
avg_block_acres	Average block size in acres
job_density_simple	Jobs per land acre simple
retail_density_simple	Retail jobs per land acre simple
job_gravity	Total jobs for every tract in the US divided by its distance from the centroid squared
retail_gravity	Retail jobs for every tract in the US divided by its distance from the centroid squared
median_commute	Median distance of commuters in the tracts using the centroid of the employment block
veh_count	Number of vehicles that were used to determine VMT from Illinois odometer reading (for appropriate Illinois tracts only)
avg_vmt	Average VMT from Illinois odometer reading (for appropriate Illinois tracts only)
avg_hh_vmt	Average VMT per household from Illinois odometer reading (for appropriate Illinois tracts only)
std_dev_vmt	Standard deviation for VMT from Illinois odometer reading (for appropriate Illinois tracts only)
area_type	'county' or 'cbsa' depending on the location of the tract
area_stfid	Census id for either county of cbsa for this tract's location
state	State FIPS codes
county	County FIPS codes

The remaining fields contain in order the model outputs given in *Table 16* for renter and owner households matching each of the eight household profiles. Note that LAIM3 includes three new fields for each household profile: income percentile in a given tract for the household profile generally, and income percentile of renter households and owner households among renters and owners in that tract, respectively. These fields have been added as a metric of how typical each household profile is income-wise. This information could help users determine how applicable the estimates for each household profile are to a particular tract or tracts of interest (see Appendix 3 for the distributions for each of these variables for each household profiles' income).

*Table 16: LAI Data Dictionary: Model Outputs for Each Household Profile*

Column Name	Description
control_hh_income_frac	Fraction of household income as defined for the household above relative to the regional median household income
control_hh_income	The household income for the geography as defined for the household above

control_hh_size	The household size for geography as defined for the household above
control_hh_commuters	The number of commuters for the geography as defined for the household above
fixes	Code if a fix was used to substitute the missing values - codes definitions in will be documented in methods doc
model_autos_per_hh_owners	The modeled autos per household for owners
model_h_cost_owners	The modeled housing costs for owners
model_pct_transit_commuters_owners	The modeled percent of transit use commuters for owners
model_vmt_per_hh_owners	The modeled household annual vehicle miles traveled (VMT) for owners
model_autos_per_hh_renters	The modeled autos per household for renters
model_h_cost_renters	The modeled housing costs for renters
model_pct_transit_commuters_renters	The modeled percent of transit use commuters for renters
model_vmt_per_hh_renters	The modeled household annual vehicle miles traveled (VMT) for renters
model_autos_per_hh	The modeled autos per household
model_h_cost	The modeled housing costs
model_pct_transit_commuters	The modeled percent of transit use commuters
model_vmt_per_hh	The modeled household annual vehicle miles traveled (VMT)
alpha	Factor used to calculate average transit cost per household
beta	Factor used to calculate average annual transit trip per household
alpha_beta_uzas	The Census Urbanized Area (UZA) used to assign alpha and beta
uza_dist	The distance to the nearest UZA used to assign alpha and beta
gas_price	The cost of gas per gallon
mpg	The national average fuel efficiency
income_bin	Indicate the income bracket for the controlled household type
auto_own_cost_owners	Auto ownership costs for owners
vmt_cost_owners	vehicle miles traveled costs for owners
transit_cost_owners	Average transit costs for owners
transit_trips_owners	The modeled household annual transit trips for owners
t_cost_owners	The modeled transportation costs for owners
t_owners	The modeled transportation costs as a percent of income for owners
h_owners	The modeled housing costs as a percent of income for owners
ht_owners	The modeled housing and transportation costs as a percent of income for owners
auto_own_cost_renters	Auto ownership costs for renters
vmt_cost_renters	vehicle miles traveled costs for renters
transit_cost_renters	Average transit costs for renters
transit_trips_renters	The modeled household annual transit trips for renters

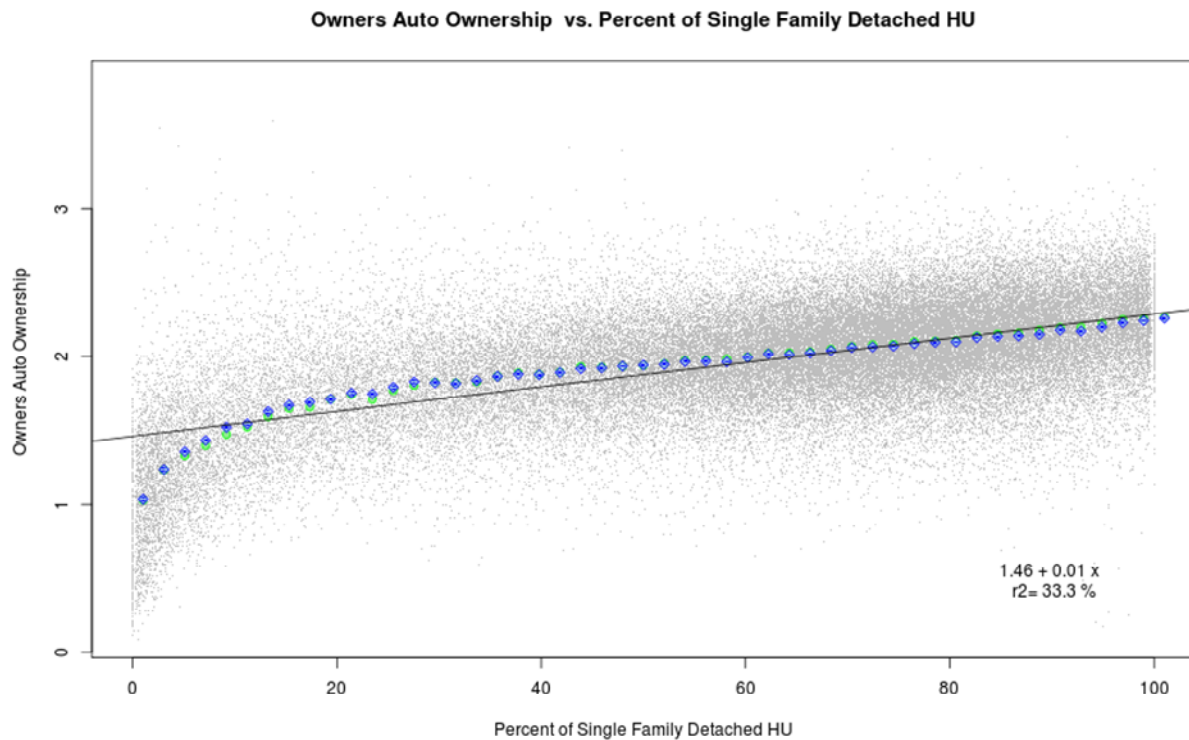
t_cost_renters	The modeled transportation costs for renters.
t_renters	The modeled transportation costs as a percent of income for renters
h_renters	The modeled housing costs as a percent of income for renters
ht_renters	The modeled housing and transportation costs as a percent of income for renters
auto_own_cost	Auto ownership costs
vmt_cost	vehicle miles traveled costs
transit_cost	Average transit costs
transit_trips	The modeled household annual transit trips
t_cost	The modeled transportation costs
t	The modeled transportation costs as a percent of income
h	The modeled housing costs as a percent of income
ht	The modeled housing and transportation costs as a percent of income
pctile_all	Income percentile in a given tract as defined for the household above
pctile_own	Income percentile for owners in a given tract as defined for the household above
pctile_rent	Income percentile for renters in a given tract as defined for the household above



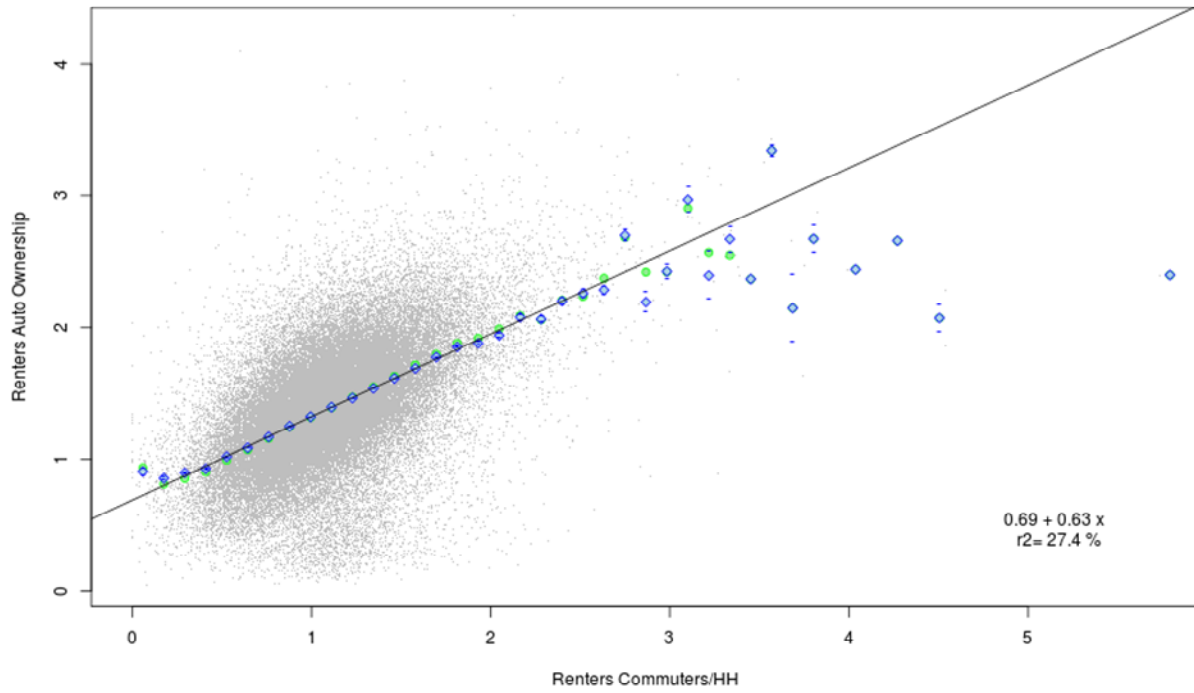
## Appendix 1: Scatter Plots of Endogenous Variables vs. an Example Exogenous Variable

The following plots show the relationships between the one of the strongest exogenous variables for each endogenous variable. Note that in each plot there are approximately 70,000 points, depending on the data suppression in the ACS. Each plot has the following features:

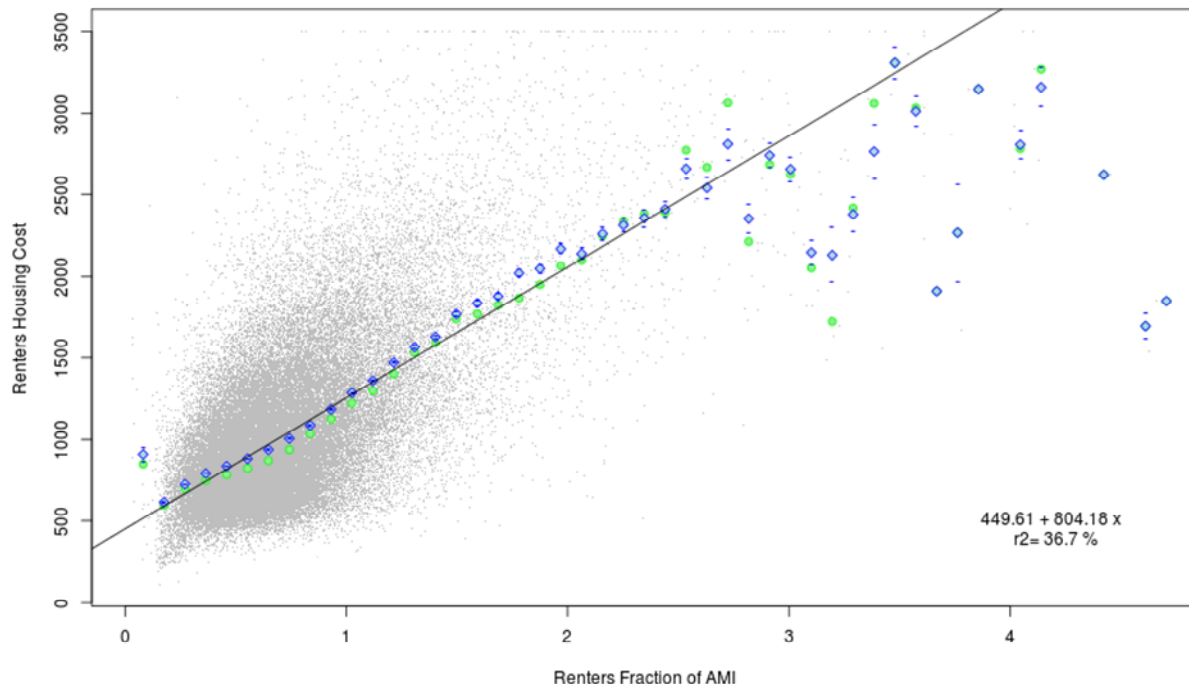
- Small grey dots – values for each census tract where there is valid data (i.e. no ACS data suppression),
- Blue diamonds with blue dashed above and below – mean value of the y variable in 50 bins of the x variable, and the blue lines represent the standard error on the mean (when there is no lines this indicates that there are only one tract 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 relationships are) and
- Text in lower right corner – the equation for the line and the  $R^2$  of the linear fit.



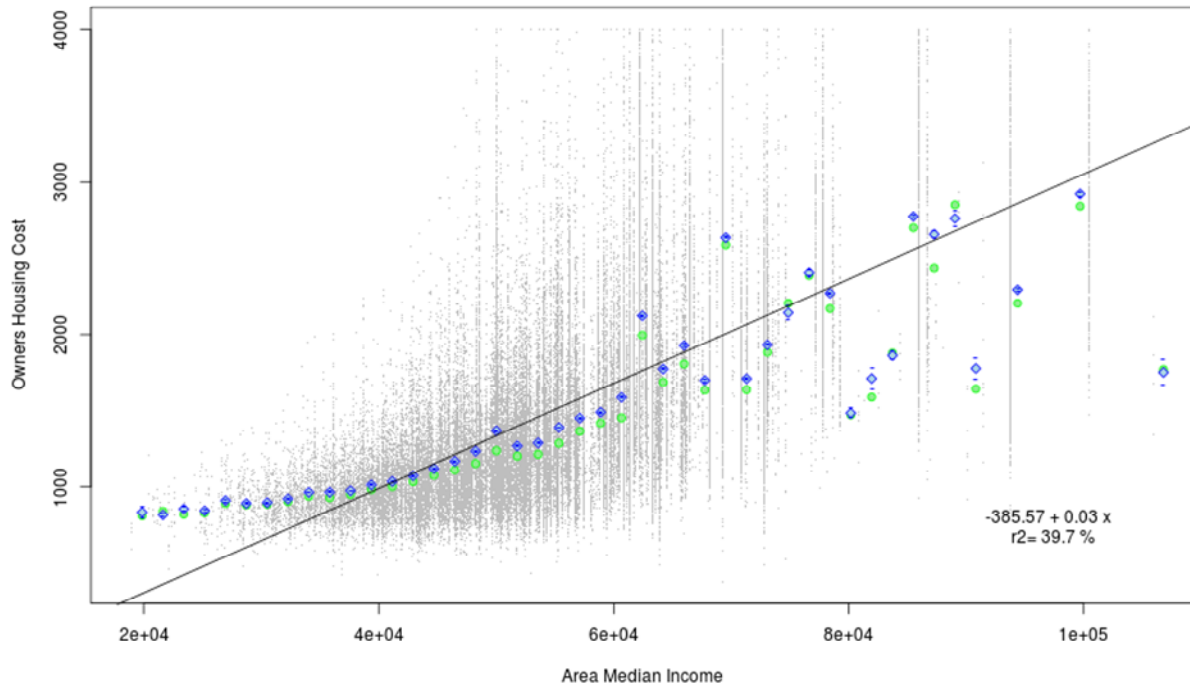
**Renters Auto Ownership vs. Renters Commuters/HH**



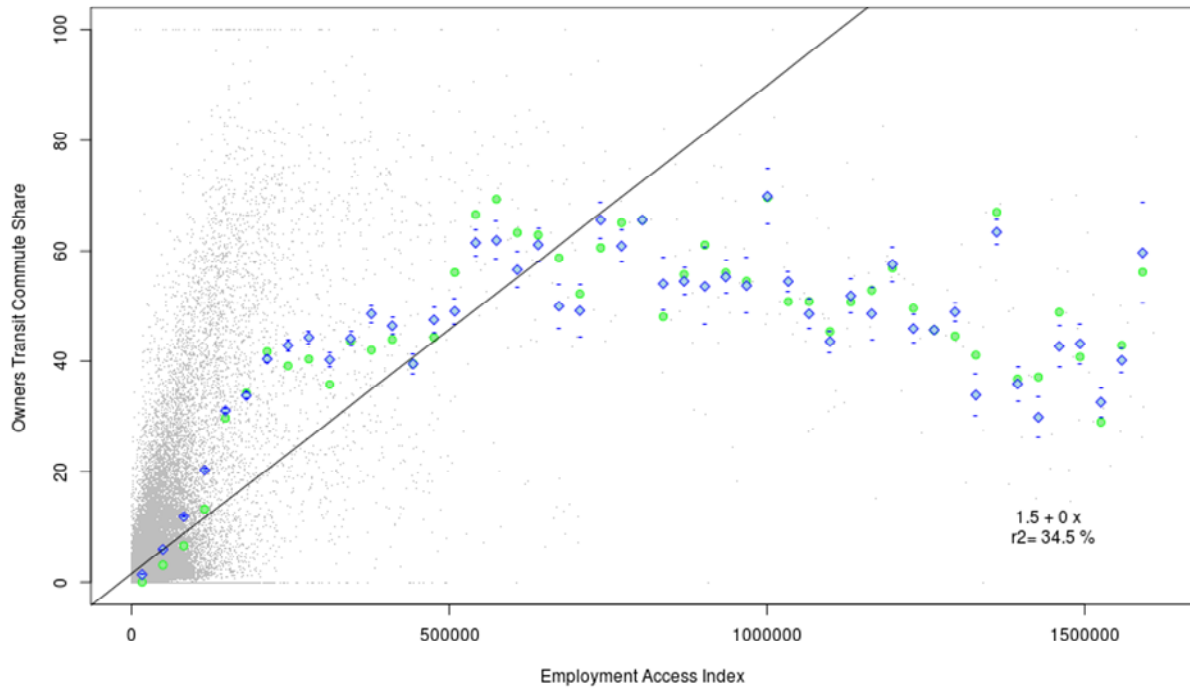
**Renters Housing Cost vs. Renters Fraction of AMI**



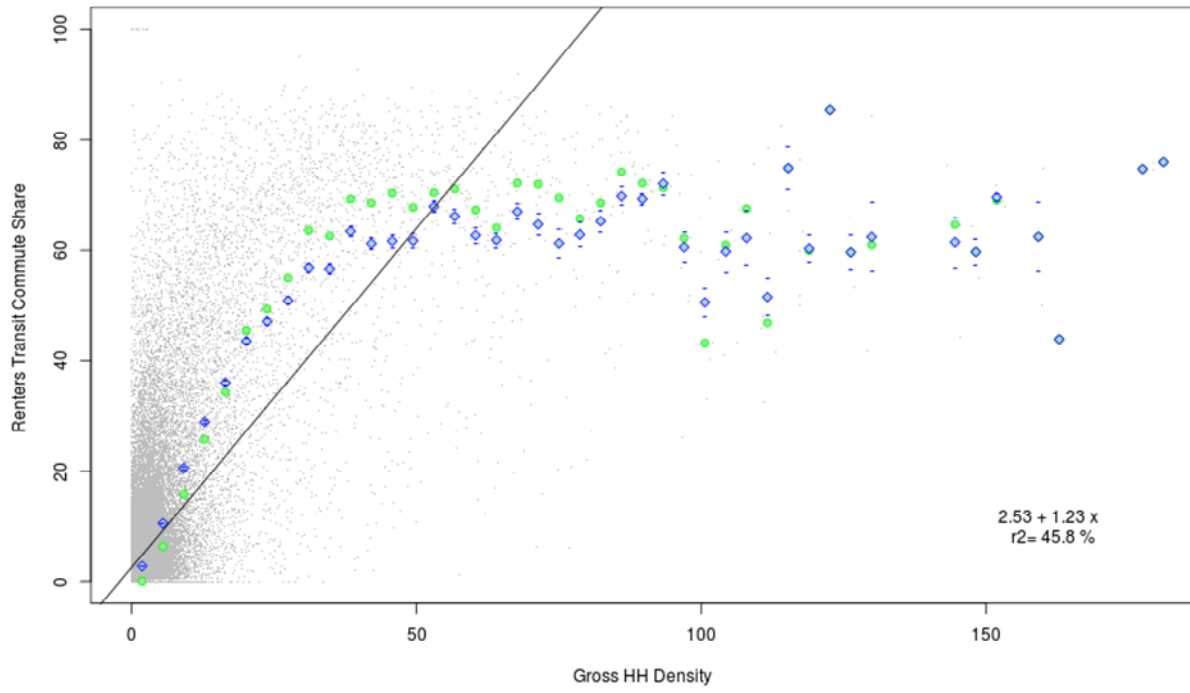
Owners Housing Cost vs. Area Median Income



Owners Transit Commute Share vs. Employment Access Index



Renters Transit Commute Share vs. Gross HH Density



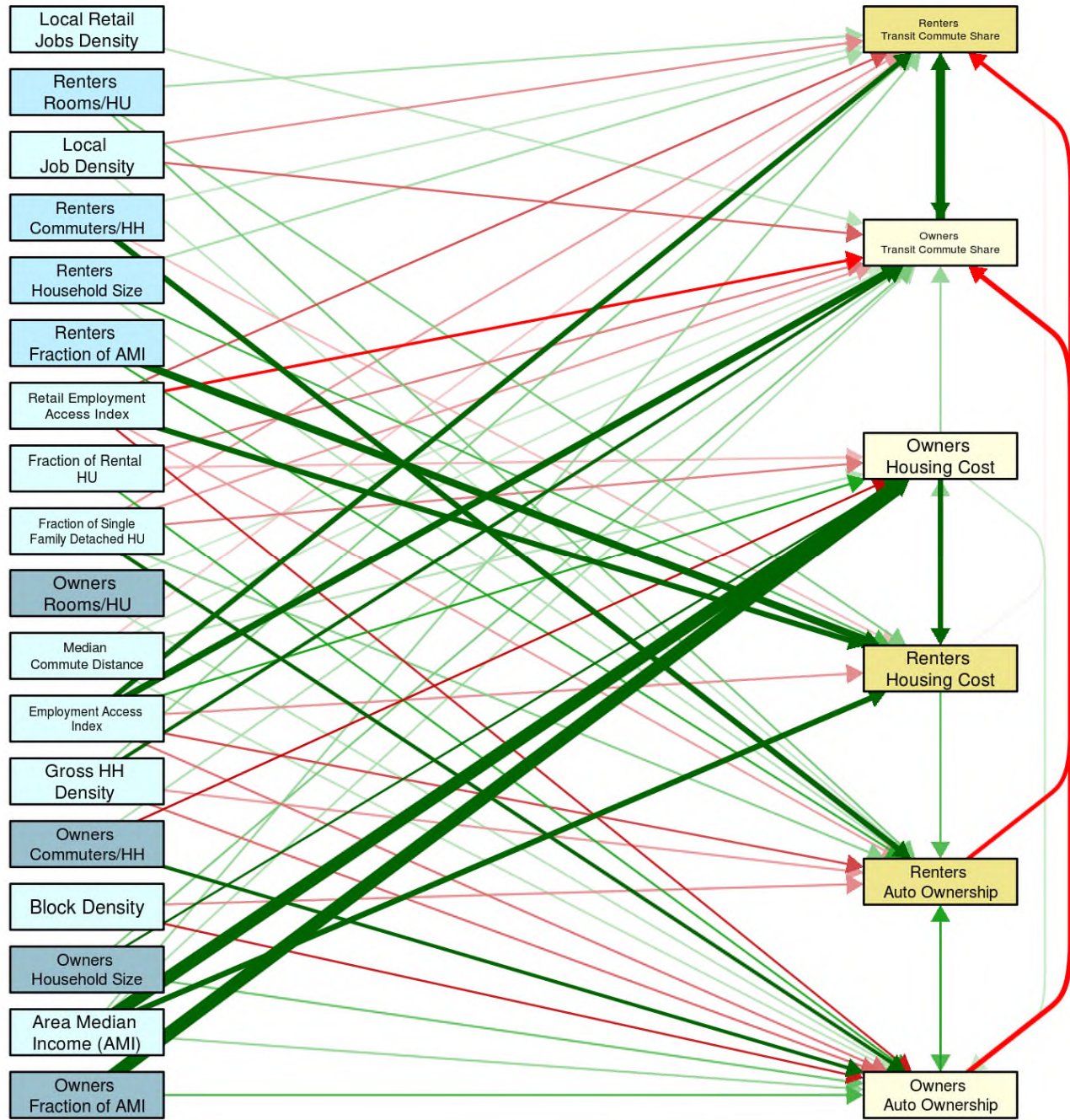
## Appendix 2: Path Diagrams

Figure 5 and Figure 6 (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 Owners Housing Cost 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 2 shows the values of the standardized variables used for LAIM3 (Figure 3 is the same diagram but with a different layout).

Figure 2: Path Diagram for SEM Model



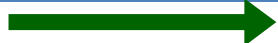





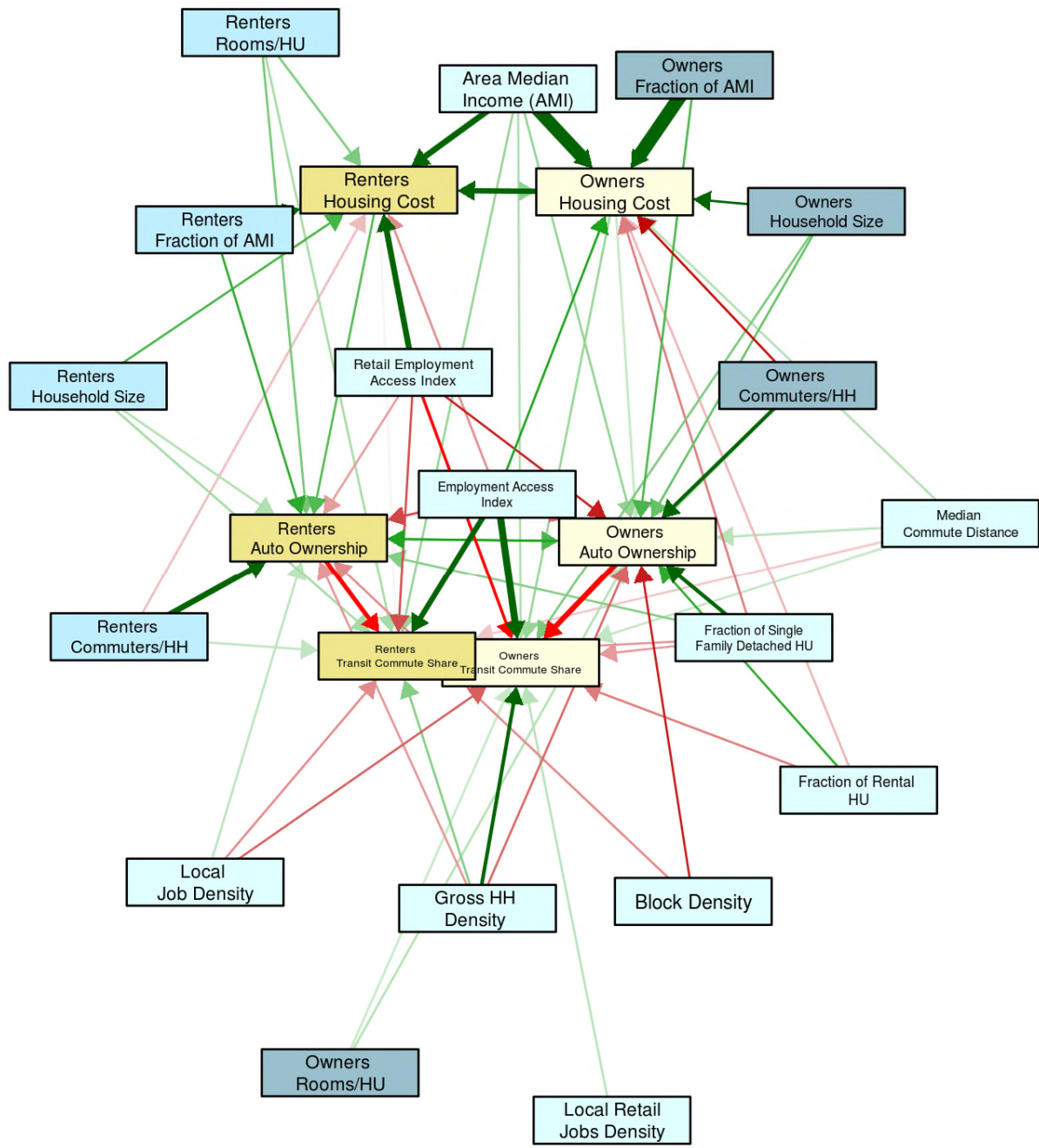
Path Diagram Key	
Line	Value
	0.50
	0.25
	0.05
	-0.05
	-0.25
	-0.30

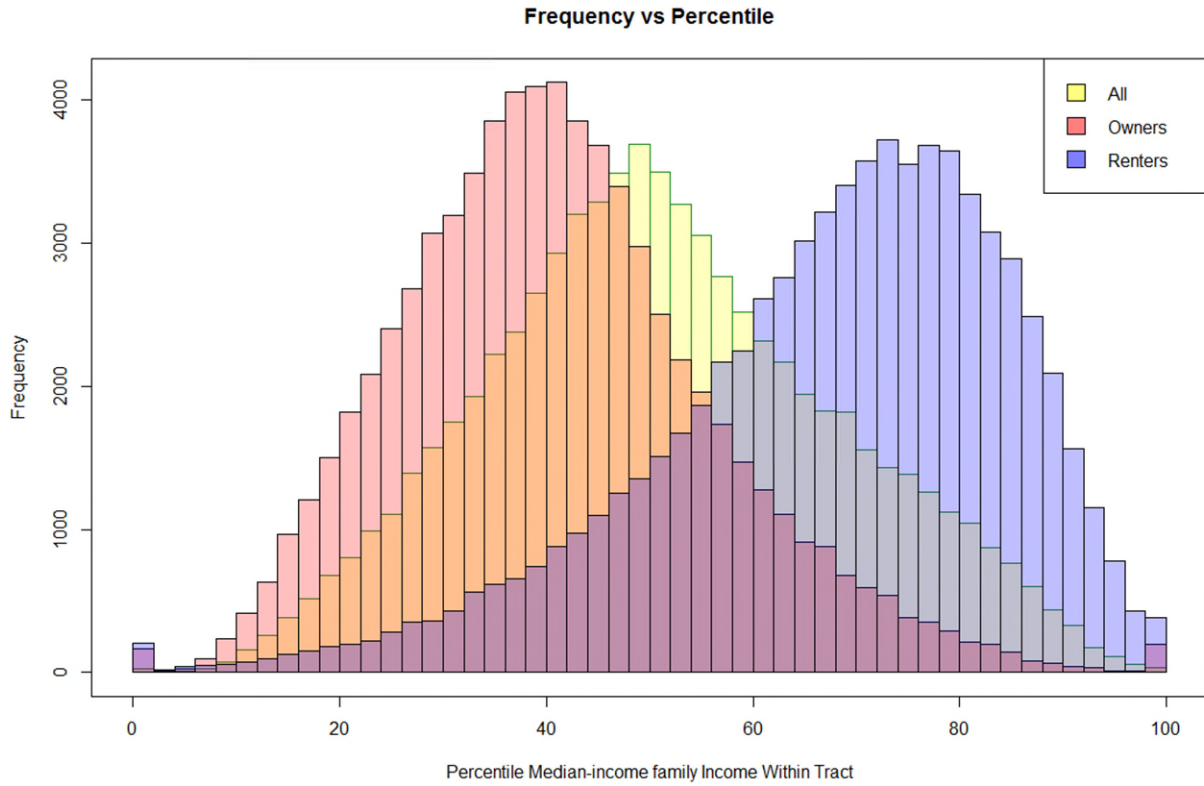
Figure 3: Path Diagram for SEM Model - Alternative Layout



Path Diagram Key	
Line	Value
	0.50
	0.25
	0.05
	-0.05
	-0.25
	-0.30

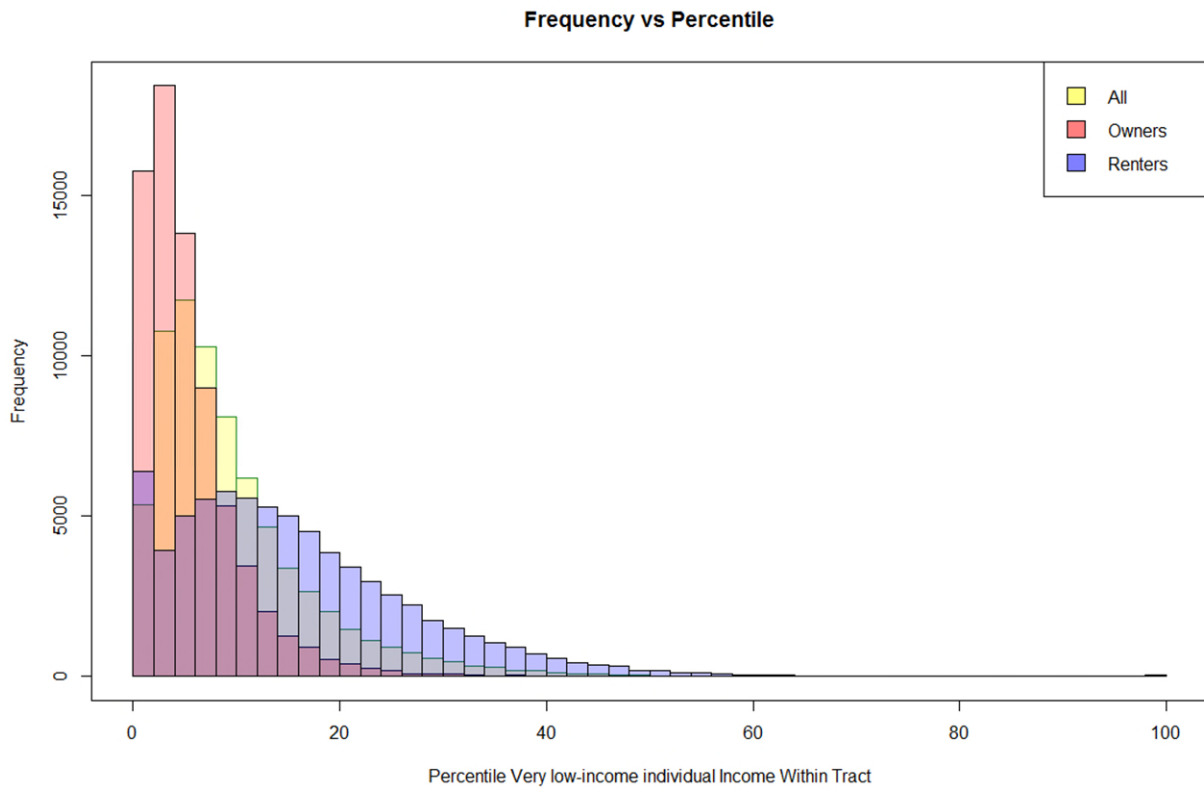
# Appendix 3: Distributions of tracts by percentile of household profile income

## Household Profile 1: Median Area Household Income

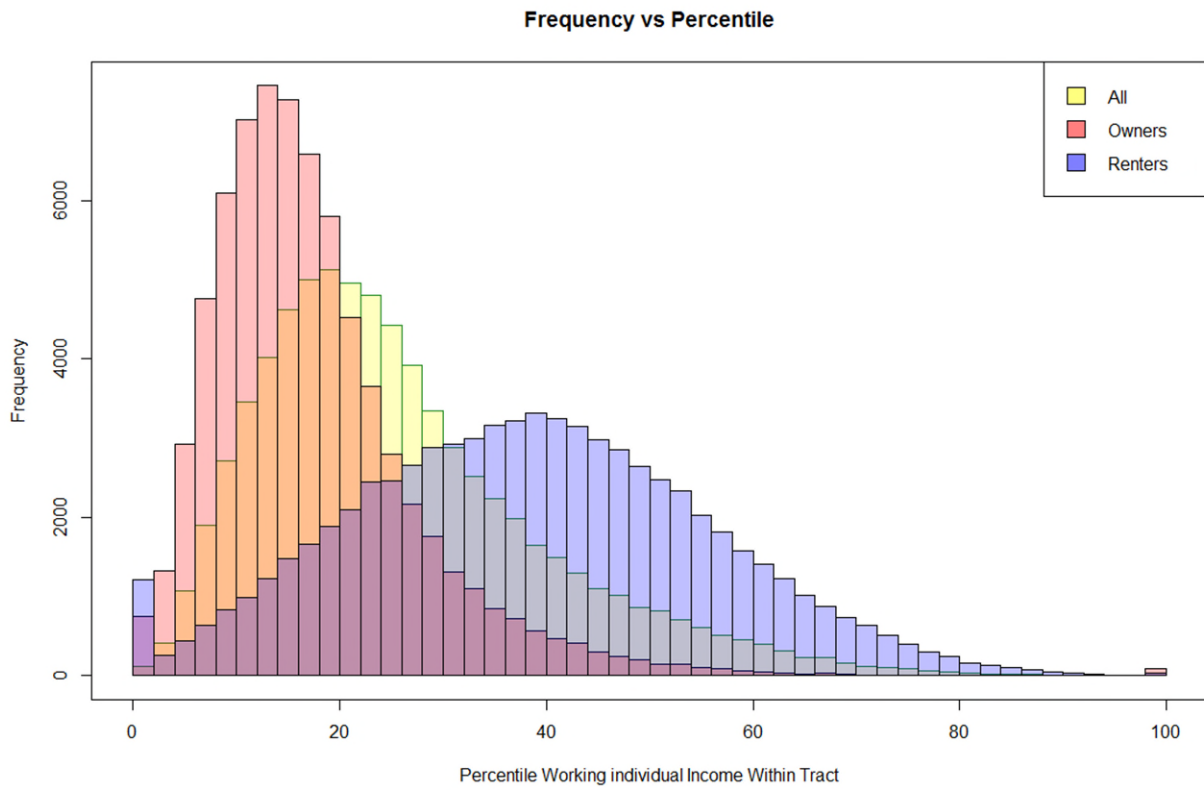




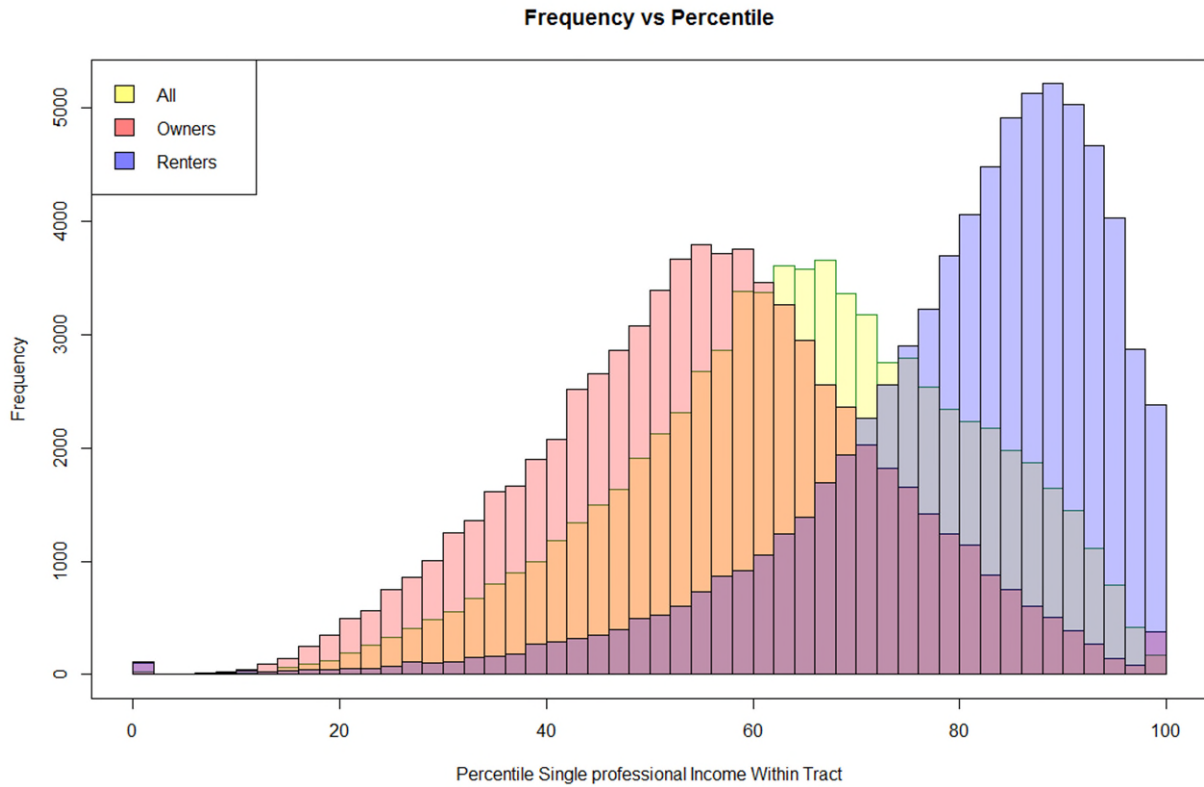
Household Profile 2: National poverty line



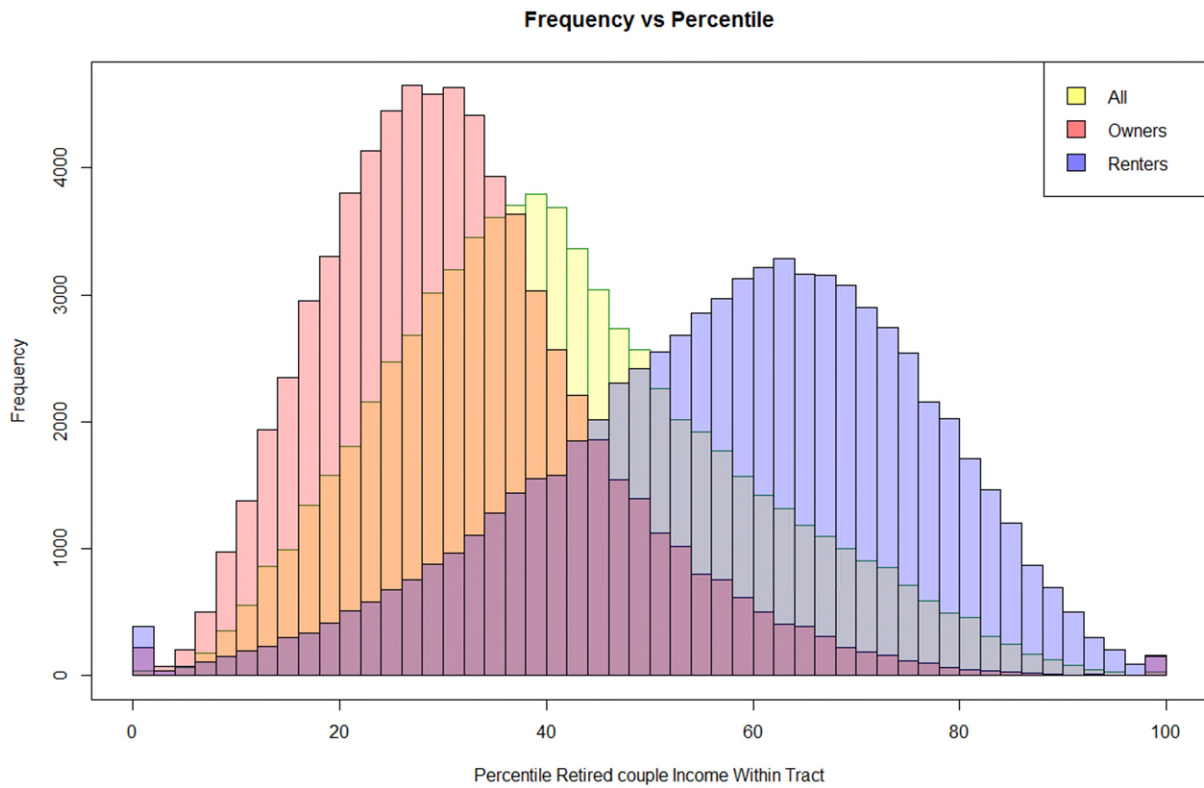
Household Profiles 3 and 6: 50% of Median Area Household Income



Household Profile 4: 135% of Median Area Household Income

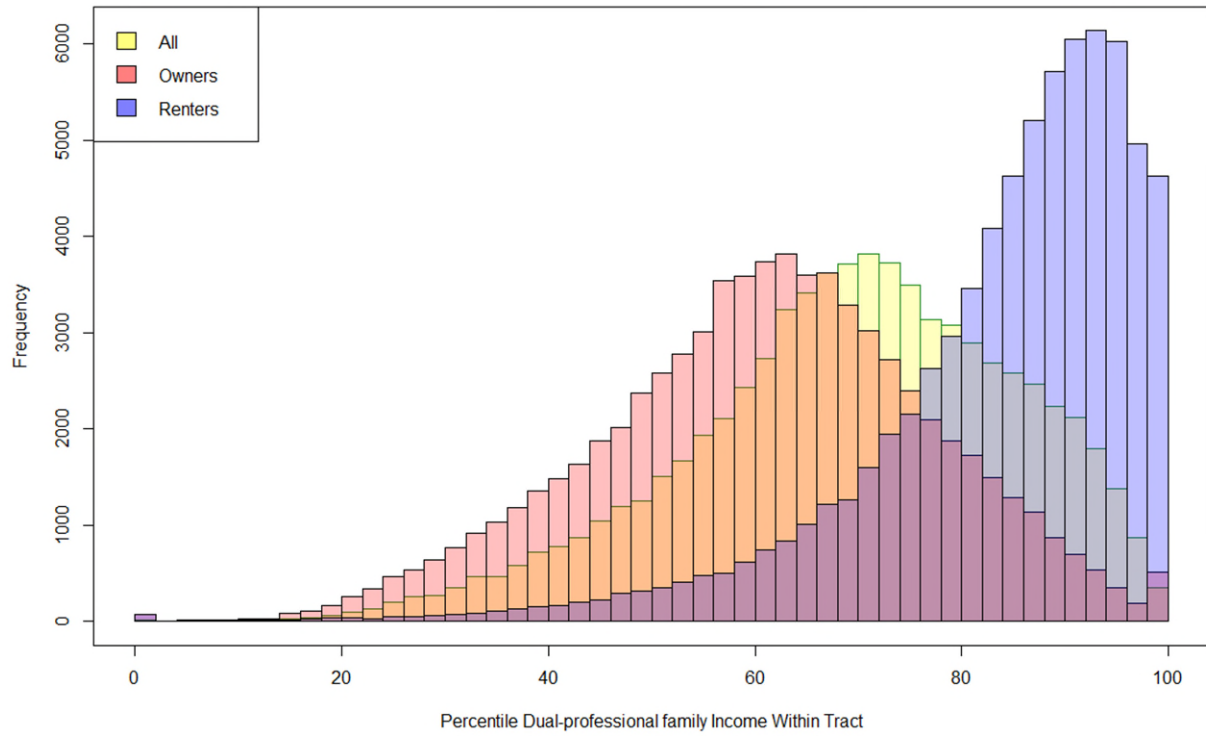


Household Profiles 5 and 7: 80% of Median Area Household Income



Household Profile 8: 150% of Median Area Household Income

Frequency vs Percentile



## Appendix 4: Version 2 Model 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. 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.

LAIM Version 2 uses both more sophisticated modeling and a refined set of variables that do a better job of representing the characteristics of the built environment relevant to housing and transportation costs.

### 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. 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.