Data and Methodology: Location Affordability Index Version 2.0

Contents

Introduction

The Location Affordability Portal (LAP), launched by the U.S. Department of Housing and Urban Development (HUD) and Department of Transportation (DOT) in November 2013, provides robust, standardized household housing and transportation cost estimates at the Census block-group level for the vast majority of the United States. These estimates are generated using the *Location Affordability Index Model* (LAIM Version 1), a combination of statistical modeling and data analysis using data from a number of federal sources. They are presented on the site in the form of two data tools: 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.

The *Location Affordability Index Model Version 2* (LAIM Version 2) represents a significant a methodological and technical advance from LAIM Version 1, in addition to updating all of the constituent data sources. 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.[1] 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.

Version 2 Model Development

During beta testing of the LAP Version 1 and subsequent discussions prior to the site's public launch[2], 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

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

[2] For complete documentation of LAIM Version 1, please see http://www.locationaffordability.info/LAPMethodsV1.pdf

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 in LAIM Version 2

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.

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

Error! Not a valid bookmark self-reference. 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 (noted by the double headed arrow).

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

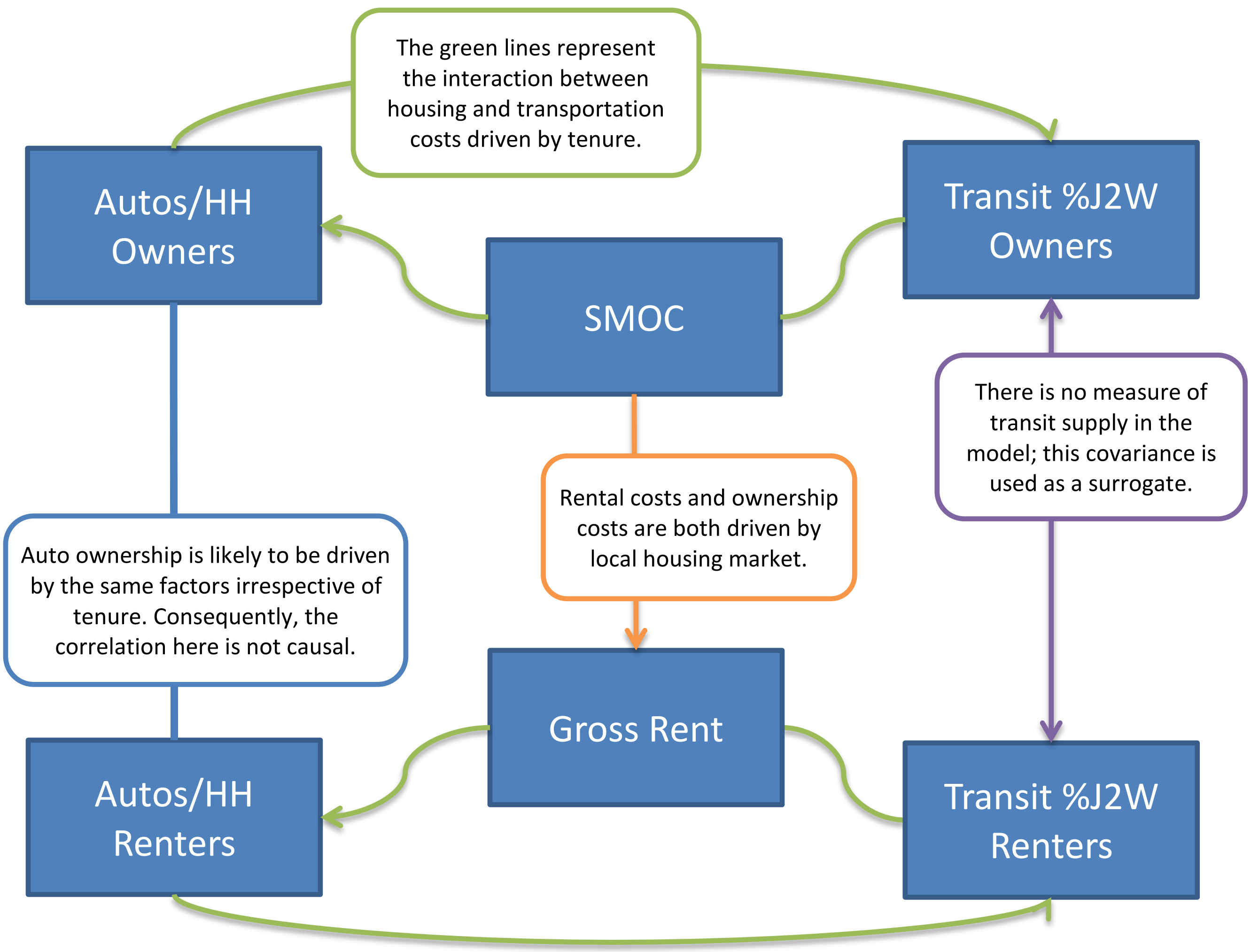


Table 1: Hypothesis of Endogenous Variable Interactions

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; therefore no explicit connection 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. Therefore there is no strong reason for a interaction and none observed.	None
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. Therefore there is no strong reason for a interaction and none 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

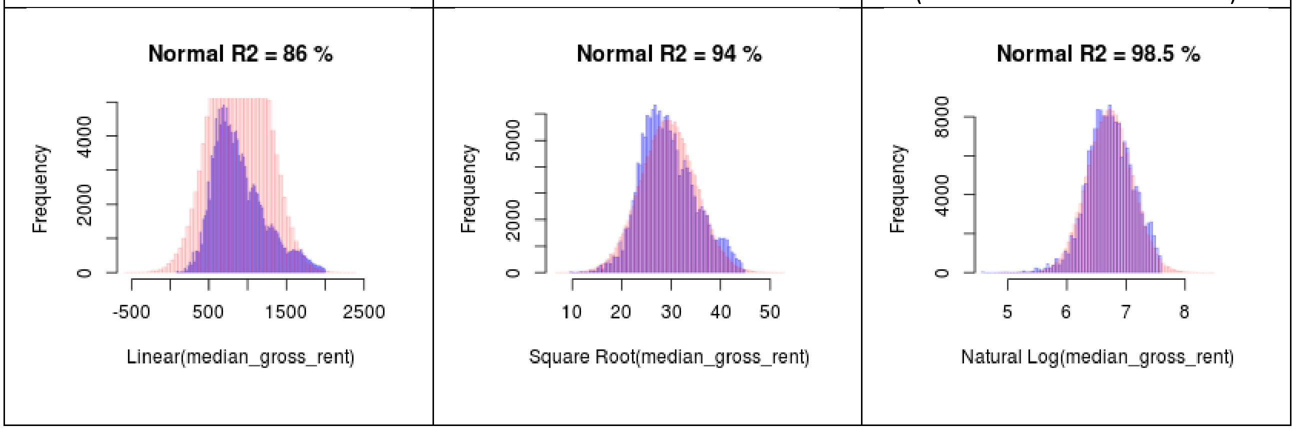
B. Variable Transformation

Once relationships between endogenous variables have been hypothesized, a preliminary model can be constructed. In LAIM Version 2, SEM variables (Table 2, next page) are transformed to allow for better fits for non-linear relationships. As shown in Figure 2 (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 graphs in Figure 2 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 R2” value is coefficient of determination of the ACS data to the normal distribution.

Figure 2: 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 2: 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[3] 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.

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

C. Variable Selection

Table 3 lists the variables used in the original LAIM but dropped from LAIM Version 2.

Table 3: 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 Street Connectivity and Walkability)
Transit Connectivity Index	Transit access as a function of transit service frequency and proximity to transit nodes, weighted by observed journey to work data	Replaced by transit commute share, a measure available for the 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 4 of this document)

Variables listed in Table 4 were added to the model based on feedback from HUD staff and a literature review of rural VMT.

Table 4: 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 centroid divided by land area of same.	Access to retail amenities.
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 4 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

D. Final Fit

The following section describes in detail final model's specification and all included variables. The structure of the model is detailed in Table 6: SEM Structure (endogenous variables are shaded) on pp. 20-22.

LAIM Version 2 Methodology

I. Geographic Level and Data Availability

LAIM Version 2 is constructed at the Census block group level using the 2012 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 2 is constructed to cover the entire United States.[4]

II. Basic Index Structure

LAIM Version 2 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, Aii. on goodness of fit measures on page 22 for further discussion). Additionally, to keep the model as simple as possible, input measures of transit access are no longer used. However since two endogenous variables are themselves measure of transit use (i.e., percent of commuters using transit for journey to work for home-owners and renters), the model works well. These revisions allow LAIM Version 2 to model housing and transportation costs by tenure for households in urban, suburban, and rural settings.

III. Data Sources

LAIM Version 2 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. 2008-2012 survey data are used in LAI Version 2.
- 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 2010 Census data, are used here.

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 the entirety 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.

[4] There are a few block groups in the United States that do not have households in them, these are not modeled.

IV. Variables

Starting with a pool of potential independent (exogenous in the SEM) variables representing all of 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 5 lists the final set of variables used in LAIM Version 2, with endogenous variables shaded.

Table 5: Overview of LAIM Version 2 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 Renters	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
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

Input	Description	Data Source
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

The following detailed descriptions of variables used for LAIM Version 2 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 A: Scatter Plots of Endogenous Variables vs. an Example Exogenous Variable show some of the relationships of the endogenous and exogenous variables.

A. 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.

i. Gross Density

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

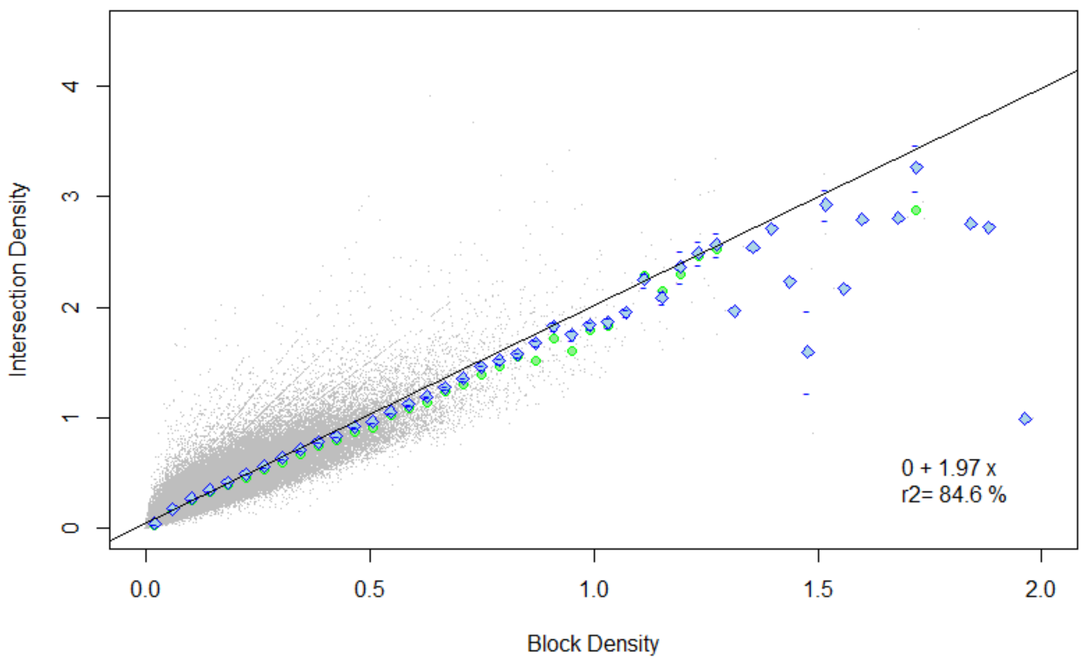
B. Street Connectivity and Walkability

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.

i. Block Density

Census TIGER/Line files are used to calculate average block density (in acres) using the number of blocks within the block group divided by the total block group land area. 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 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 LAIM Version 2. Figure 3, which shows the correlation between the measures, illustrates just how collinear these two measures are.

Figure 3: Intersection Density (Intersections per Acre) versus Block Density (Blocks per Acre) for all U.S. Census Block Groups



C. 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 group level. These data are currently unavailable in Massachusetts.[5]

[5] Using Massachusetts ES202 database query tool (http://lmi2.detma.org/lmi/lmi_es_a.asp) the employment by county was obtained for 2010. Using a constant share method from the 2000 CTPP employment data at the block group level, an estimate of 2010 employment was made for every block group in Massachusetts.

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.

i. 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 block group. 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 block group. Because a gravity model enables consideration of jobs both directly in and adjacent to a given block group, 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 block group

n = total number of Census block groups

p_i = number of jobs in the i^{th} Census block group

r_i = distance (in miles) from the center of the given Census block group to the center of the i^{th} Census block group

As jobs get farther away from the 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.

ii. Retail Employment 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)

iii. Median Commute Distance

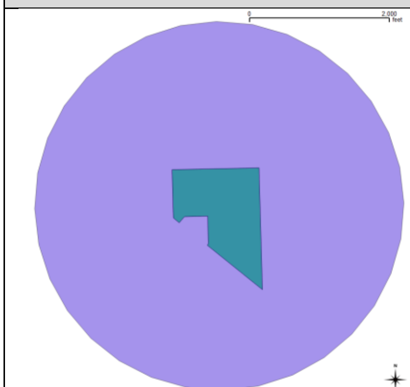
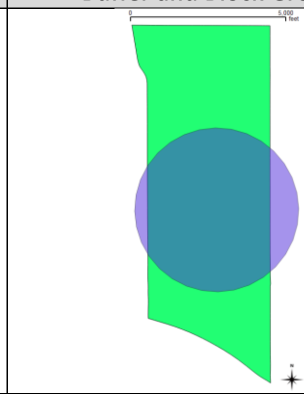
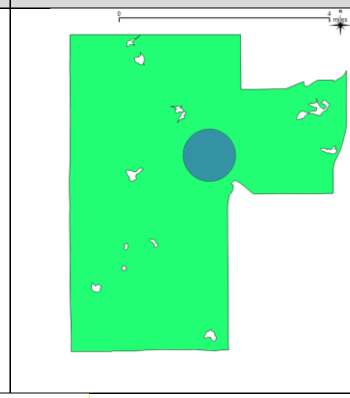
Median commute distance is calculated using LODES data. Median distances are calculated for each Census block using Euclidean (as the crow flies) distances between the origin and destination Census blocks. Block values are then sorted by distance to obtain the median value for the block group of interest.

iv. Local Job Density

Three different steps are considered to determine local job density, all of which use a half-mile buffer around the centroid of each block group (the centroid, in this case, is defined by the average of the block centroids weighted by households from 2010 Census). Using LODES data, the total number of jobs in the buffer is calculated and divided by the land area. The jobs and land is derived in one of three ways depending on the size of the block group. Figure 4 on the following page illustrates three possible scenarios:

a. If the border of the block group is completely within the half-mile buffer zone, the half-mile buffer value is used;
b. If the union[6] of the half-mile buffer and block group polygons are about the same, the value is determined by the polygon; or
c. If the half-mile buffer is completely inside the block group, the block group value is used.

Figure 4: Three Scenarios Considered for Local Employment Density Measures

Jobs in ½ Mile Buffer/ Land Area in ½ Mile Buffer	Jobs in Union of Buffer and Block Group/Land Area in Union of Buffer and Block Group	Jobs in Block Group/ Land Area in Block Group
		
	Half Mile Buffer Centered at Household Centroid Census Block Group	

v. Local Retail Density

The same three steps used to determine local job density are used for local retail density. After constructing a half-mile buffer around the centroid of each block group, LODES data is used to calculate the total number of retails jobs in the buffer, which is then divided by the land area.

a. If the border of the block group is completely within the half-mile buffer zone, the half-mile buffer value is used;
b. If the union[7] of the half-mile buffer and block group polygons are about the same, the value is determined by the polygon; or
c. If the half-mile buffer is completely inside the block group, the block group value is used.

Again, Figure 4 illustrates the three possible scenarios.

[6] "Union" is a GIS term which refers to the merging of two polygons into one. All three steps used to determine employment and retail density use the "union" of two polygons: the half-mile buffer and the block group.

[7] "Union" is a GIS term which refers to the merging of two polygons into one. All three steps used to determine employment and retail density use the "union" of two polygons: the half-mile buffer and the block group.

D. Housing Characteristics

Characteristics of the housing stock and tenure have been found to have an effect on 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 2012 ACS 5-year estimates serve as the data source for variables pertaining to housing characteristics.

i. 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.

ii. 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.

iii. 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.

iv. 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.

E. Household Characteristics

The 2012 ACS 5-year estimates serve as the primary data source for variables pertaining to household characteristics.

i. Area Median Income

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

ii. Fraction of Area Median Income Owners

Fraction of area median income for owners is calculated as the ratio of median income for owners at the block group level to the Area Median Income (see paragraph E.i.). In cases where the block group 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.

iii. Fraction of Area Median Income Renters

Fraction of area median income for renters is calculated as the ratio of median income for renters at the block group level to the Area Median Income (see paragraph E.i.). In cases where the block group 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.

iv. 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 block group population in owner occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

v. 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 block group population in renter occupied households is suppressed the value for the tract is used in running the model but not for calibrating the model.

vi. 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 block group population in owner occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

vii. 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 block group population in renter occupied households is suppressed, the value for the tract is used in running the model but not for calibrating the model.

F. Housing Costs

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

i. 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.

ii. 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 Transportation Behavior

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

i. Autos per Household Owners

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

ii. Autos per Household Renters

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

iii. Percent Transit Journey to Work Owners

As no direct measure of transit use is available at the block group 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.

iv. Percent Transit Journey to Work Renters

As no direct measure of transit use is available at the block group 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.

V. Model Structure and Formula

A. Simultaneous Equations Model

As previously mentioned, the SEM used in LAIM Version 2 consists of six nested equations, each drawing from a pool of 18 exogenous variables, that predict six interrelated endogenous variables.

i. SEM Structure

Table 6 (following page) shows the structure of the SEM model used in LAIM Version 2, organized by the six nested equations for the model’s endogenous variables (which are shaded and bolded). All endogenous variables appearing as exogenous variables in other nested equations are shaded as well.

Table 6: SEM Structure (endogenous variables are shaded)

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
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

Variables	Estimate	Std. Error	Z-Value
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			
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

Variables	Estimate	Std. Error	Z-Value
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

See Appendix B: for a path diagram that illustrates theses coefficients. Table 7 (following page) 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 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.

ii. Evaluation Metrics

The complexity of SEMs has resulted in a range of metrics to assess the model goodness of fit. For the particular SEM employed in LAIM Version 2, 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. The rule of thumb that Kline reports is that an "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 0.053 whose 90% confidence interval ranges from 0.052 to 0.054.

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. 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 model has a CFI of 0.970.
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. The rule of thumb Kline reports is that "values of the SRMR less than 0.10 are generally considered favorable." The model has an SRMR of 0.013.

By achieving these three robust measures, the SEM model used for LAIM Version 2 is shown to be a good statistical model.

B. Vehicle Miles Traveled

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 households that drive their autos, calculated from odometer readings from the Chicago and St. Louis metro areas for 2008 through 2010, obtained from the Illinois Environmental Protection Agency. Two odometer readings—for 2008 and 2010—were matched for over 660,000 vehicles using vehicle identification numbers (VIN) to obtain data for VMT during that period.

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. 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 assigning them to Census block groups using ZIP+4™ geographical identifications. Automobiles were matched using their VIN and the total distance driven was determined over the time period between inspections. The resulting analysis showed that the ratio of the average VMT predicted by the LAI VMT model to average ANNMILES by census region was 1.08,[8] suggesting that the LAI VMT model slightly underestimates auto usage nationwide. Previous analysis suggests that most of this discrepancy is 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 additional factor of eight percent. To reduce any bias in the model, this factor is estimated by comparing the 2009 National Household Travel Survey (NHTS) to the modeled value of the NHTS field ANNMILES, which is the self-reported miles driven for each auto.

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

In both versions of the LAIM, 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., household density and household income are separate inputs; the combination of the two are also used as inputs. The independent variables used in the regression are essentially the same as the exogenous variables for SEM[9] and were linearized in the same way as in the SEM analysis. The choice of transformation was made to optimize the distribution of the variables such that the distribution of the transformed variable was the most Gaussian or Normal. All Census block groups covered by the Illinois odometer data were used for the auto use regression.

Additionally, because there is an inherent spatial autocorrelation for the dependent variables, a robust variance calculation is employed to estimate the statistical significance of the regression coefficients. The method for estimating the error on the coefficients uses geographical clustering. Three natural geographical clustering definitions were tested: state, county 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 CBSA clustering was employed.

There is a high probability that the independent variables are multi-collinear. To eliminate as much of this as possible, the variance inflation factor (VIF)[10] was examined. 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 10,000 to begin with, and drop perceptibly as highly multi-collinear coefficients were excluded.

Table 8 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 9: Regression Coefficients for VMT Model on the next page. Note that there is no significant relationship with median rooms per housing unit (also retail gravity and local job density) this result leads to a need of only one model run per household type since there is not dependence on tenure.

[9] The one difference is that this 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.
[10] For a definition of VIF see http://en.wikipedia.org/wiki/Variance_inflation_factor .

Table 8: Independent Variables Used in VMT Regression

Variable Name	Linear Transformation	Linearized Variable Name	Number of Times Used in Combination
Area Income Fraction	Natural Log	area_income_frac	3
Area Median Income	Natural Log	area_median_hh_income	1
Median Journey to Work Miles	Natural Log	avg_d	2
Avg HH Size	Natural Log	avg_hh_size	2
Block Density	Square Root	block_density	1
Commuters/HH	None	commuters_per_hh	3
Employment Access	None	emp_gravity	1
Fraction Rental Units	Square Root	frac_renters	2
Gross HH Density	Square Root	gross_hh_density	2
Local Job Density	Square Root	le_jobs_total_per_acre	0
Local Retail Jobs per acre	Square Root	le_job_type_07_per_acre	2
Median Room/HU	None	median_number_rooms	0
Fraction Single Detached HU	None	pct_hu_1_detached	1
Retail Gravity	Natural	retail_gravity	0

Table 9: Regression Coefficients for VMT Model

Variable	Value	Standard Error	VIF
Intercept	6227.041	569.631	0.000
avg_hh_size*pct_hu_1_detached	7.834	0.635	3.492
$emp\_gravity^2$	-54.489	3.696	4.367
area_income_frac*avg_hh_size	810.671	80.833	3.249
area_median_hh_income*commuters_per_hh	1226.343	30.785	2.310
area_income_frac*gross_hh_density	-831.642	61.461	2.914
avg_d*frac_renters	-1978.234	129.015	3.945
avg_d^2	457.996	32.741	3.295
$gross\_hh\_density^2$	-135.155	12.056	1.645
block_density*commuters_per_hh	-3116.456	270.823	3.005
area_income_frac*frac_renters	2620.637	512.846	3.804
commuters_per_hh*le_job_type_07_per_acre	-857.818	138.784	4.661
$le\_job\_type\_07\_per\_acre^2$	191.503	37.079	3.465

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

To hone in on 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.

I. Modeling Transportation Behaviors and Housing Costs

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 types are enumerated in Table 10. 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 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 (see Table 10). It was run for both owner and renter tenure for each type.

Table 10: LAI Household Types

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
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 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 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 Table 11).
2. The VMT model was run for each household type, irrespective of tenure.
3. The model SMOC was evaluated and adjusted using the following criteria: if the value was less than the 10 percentile, overwrite the modeled value with the 10 percentile value; if over the 90 percentile, overwrite modeled value with the 90 percentile value.
4. The modeled gross rent was evaluated and modified in the same way as step 3
5. 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 2012 dollars from the 2010 calculations.

6. Put costs together with the ratio of each household type income and integrate into the database.

Table 11: Household Variables used in SEM

Modeled Variables	Owner Household Variables[11]	Renter Household Variables[12]
• Autos/HH Owners • SMOC • Transit %J2W Owners	Values from Table 10	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 10

Some notable differences between LAI Version 1 and LAI Version 2 resulting from 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[13] 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.

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.**

   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,

[11] Household Income Owners, Household Size Owners, and Commuters per Household Owners
[12] Household Income Renters, Household Size Renters, and Commuters per Household Renters
[13] Econsult Solutions conducted a third-party review of LAIM Version 1 to assess the validity of the model and provide recommendations for potential improvements.

making it possible to develop only one model for all census block groups (both urban and rural) for the entire country.

II. Transportation Cost Calculation

As discussed, LAIM Version 2 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. This operation is performed for the estimates generated for each of the eight household types.

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 Version 2. Research conducted by Diane Schanzenbach, PhD and Leslie McGranahan PhD, 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 ($0-$20,000; $20,000-$40,000; $40,000-$60,000; $60,000-$100,000; and, $100,000 and above) and applied to the modeled autos per household and annual VMT for the appropriate income range. LAI Version 2 uses an additional inflation factor of 1.052913[14] to adjust to 2012 dollars.

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

- Average annual service flow value[15] 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 12: 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-$20,000)	$2,396	$73	$657.3	$400.8	$1,182.0	1.4	0.34
2 ($20,000-$40,000)	$2,478	$133	$732.0	$421.1	$1,369.5	1.6	0.31
3 ($40,000-$60,000)	$2,586	$182	$755.6	$458.8	$1,494.2	1.9	0.31
4 ($60,000-$100,000)	$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

[14] http://www.bls.gov/data/inflation_calculator.htm
[15] Service flow is the average annual dollar amount of depreciation the vehicle has lost over the time of ownership.

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_{sf} = Per vehicle service flow cost from Table 12 (1) – for the appropriate income group
V_{fc} = Per vehicle finance charge from Table 12 (2) – for the appropriate income group
V_{fixed} = Per vehicle (fixed) ownership cost from Table 12 (3) – for the 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 2008)[16]
R = the Average Ratio drivability to fuel cost from Table 12 (7) – for the appropriate income group

B. Transit Use Costs

Transit cost data were obtained from the 2010 National Transit Database (NTD).[17] Specifically, we looked at directly operated and purchased transportation revenue as reported by each transit agency in the database. [18] Most transit agencies serve only one CBSA, but there are a number of larger systems that serve multiple CBSAs, which requires their revenue to allocated 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.

To illustrated, 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:

$$CBSA_1 = (1000*1000)/(1000*1000 + 200*85) = 98 \text{ percent}$$

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

[16] U.S. Department of Energy, Energy Information Administration. "Petrolium & Other Liquids." Accessed from http://www.eia.gov/petroleum/gasdiesel/.
[17] http://www.ntdprogram.gov/ntdprogram/datbase/2010_database/NTDdatabase.htm
[18] Demand response revenue is not factored into this analysis.

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 households and the percent of journey to work trips.[19]

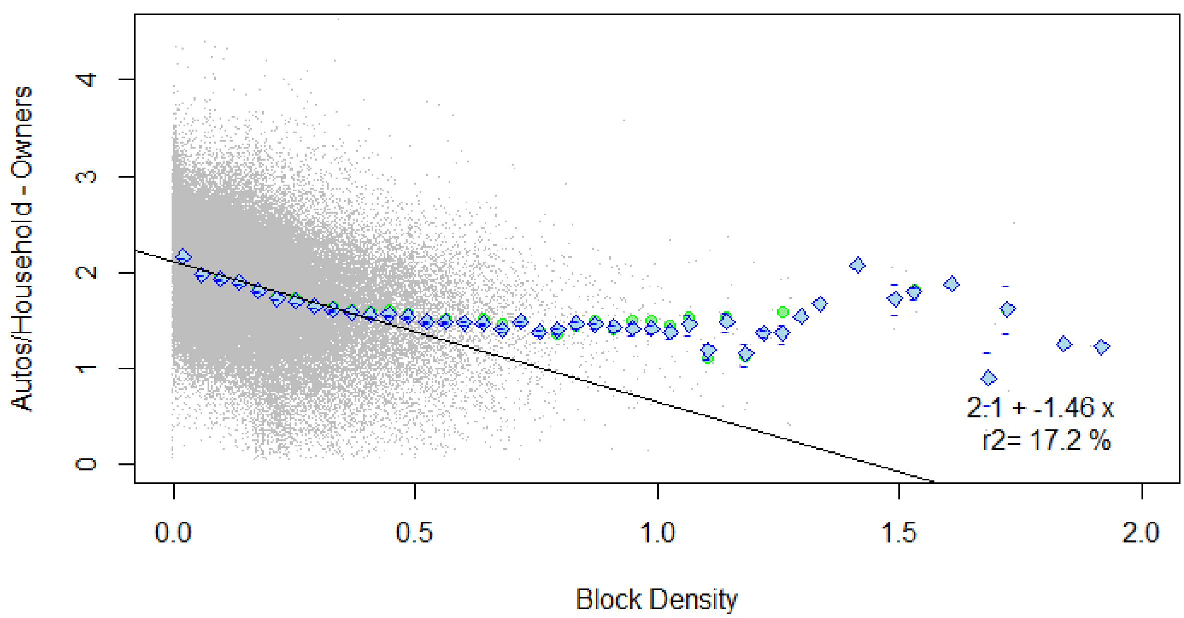
There are a number of metropolitan areas without sufficient information on transit stop locations and/or no revenue listed in the NTD. The average from the allocation calculation described in the previous paragraph is used for these metropolitan areas. The average transit costs are then allocated to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

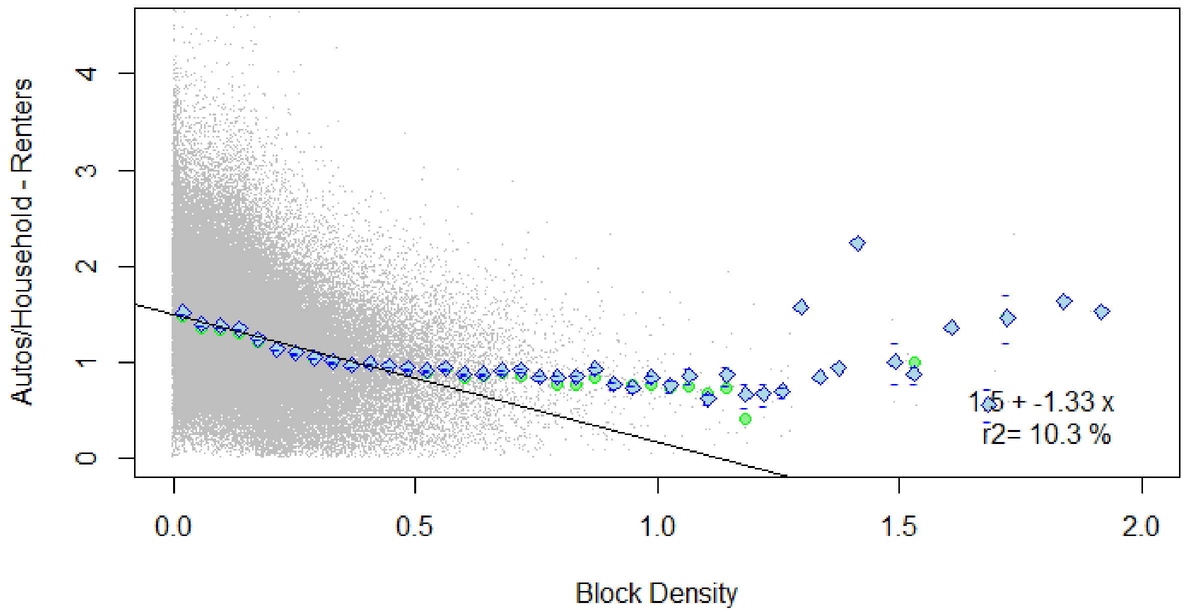
[19] This normalization method assumes that the transit use for the journey to work is a good surrogate for overall transit use.

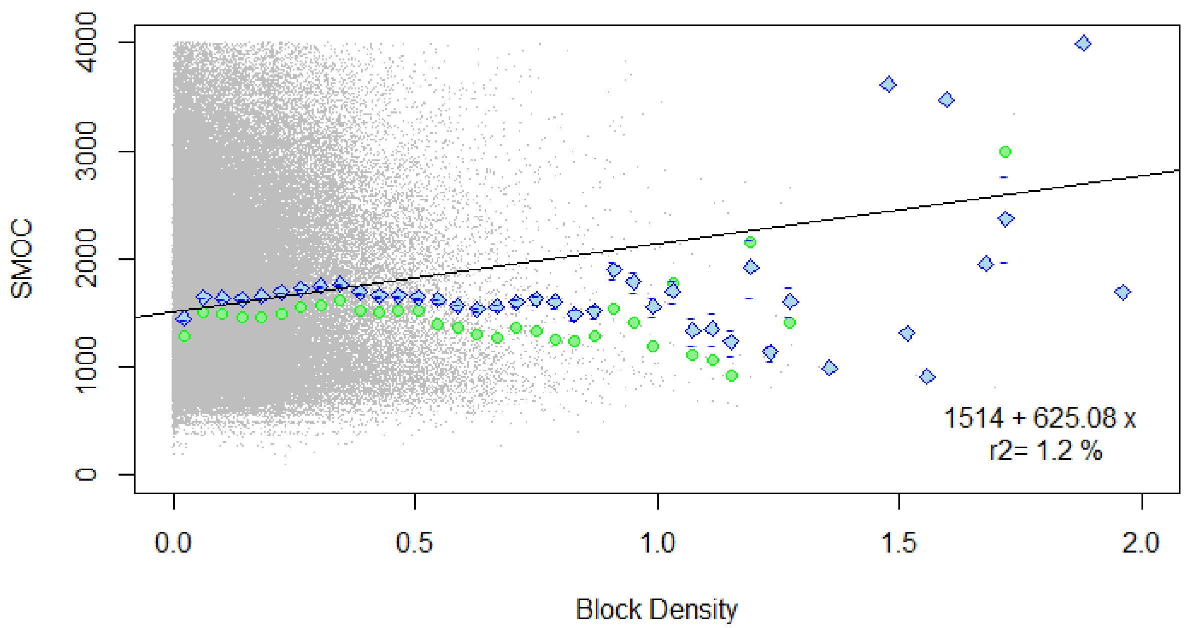
Appendix A: Scatter Plots of Endogenous Variables vs. an Example Exogenous Variable

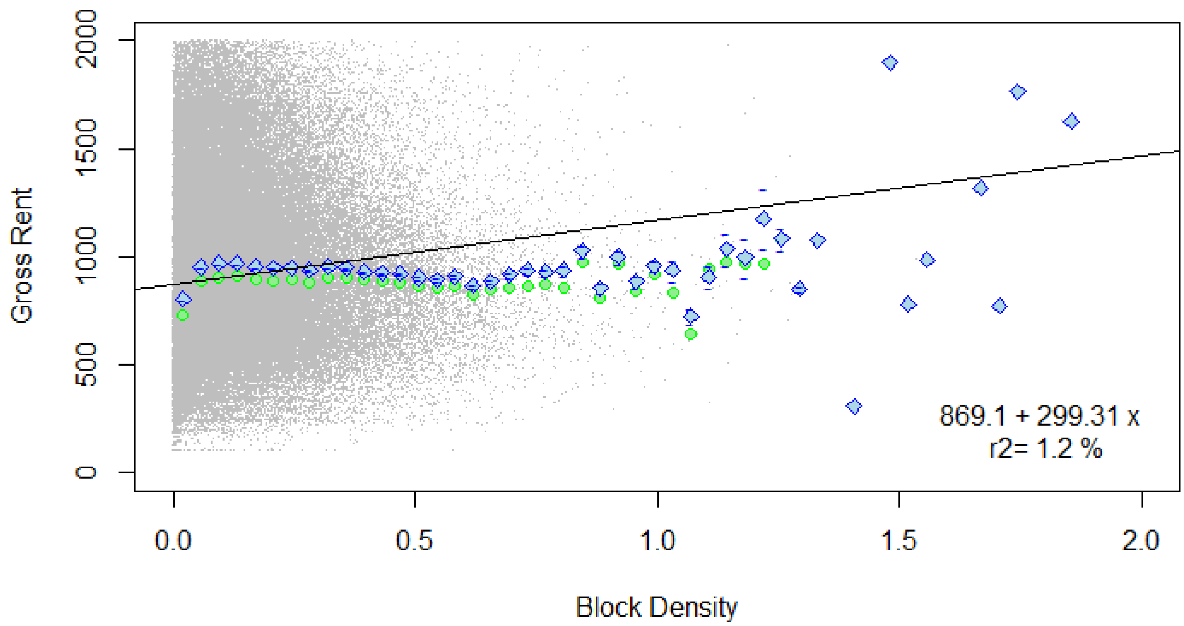
The following plots show the relationships between some of the exogenous variables and the endogenous variables. Note that in each plot there are approximately 200,000 points, depending on the data suppression in the ACS. Each plot has the following features:

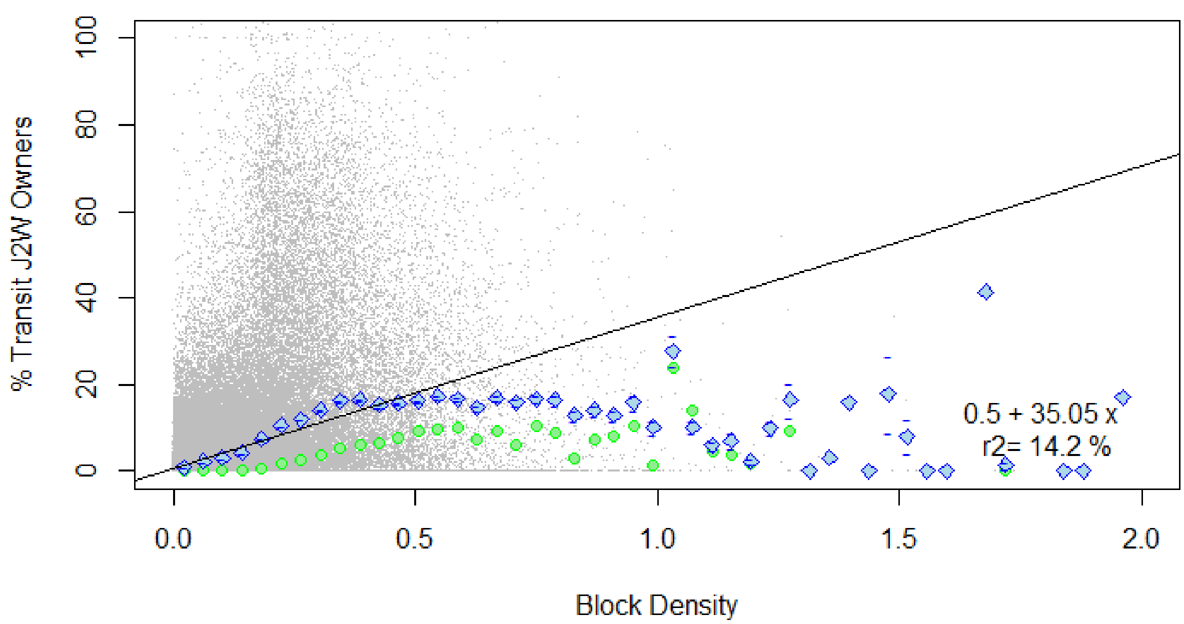
- Small grey dots – values for each census block group where there is valid data (i.e. no ACS data suppression),
- Blue diamonds with blue dashed above and below – mean value of the y variable in 50 bins of the x variable, and the blue lines represent the standard error on the mean (when there is no 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.

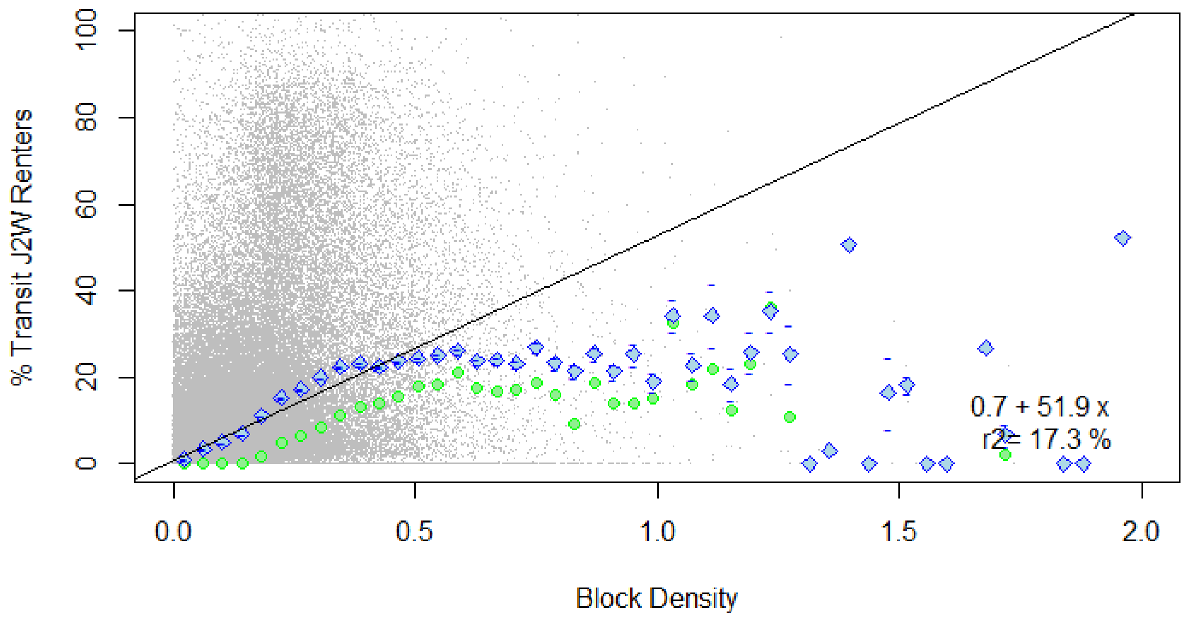



Autos/Household - Renters
0
1
2
3
4
0.0
0.5
1.0
1.5
2.0
Block Density
1.5 + -1.33 x
r2= 10.3 %


SMOC
0
1000
2000
3000
4000
0.0
0.5
1.0
1.5
2.0
Block Density
1514 + 625.08 x
r2= 1.2 %


Gross Rent
0
500
1000
1500
2000
0.0
0.5
1.0
1.5
2.0
Block Density
869.1 + 299.31 x
r2= 1.2 %


% Transit J2W Owners
0
20
40
60
80
100
0.0
0.5
1.0
1.5
2.0
Block Density
0.5 + 35.05 x
r2= 14.2 %



% Transit J2W Renters
Block Density
0.7 + 51.9 x
r2= 17.3 %
0
20
40
60
80
100
0.0
0.5
1.0
1.5
2.0

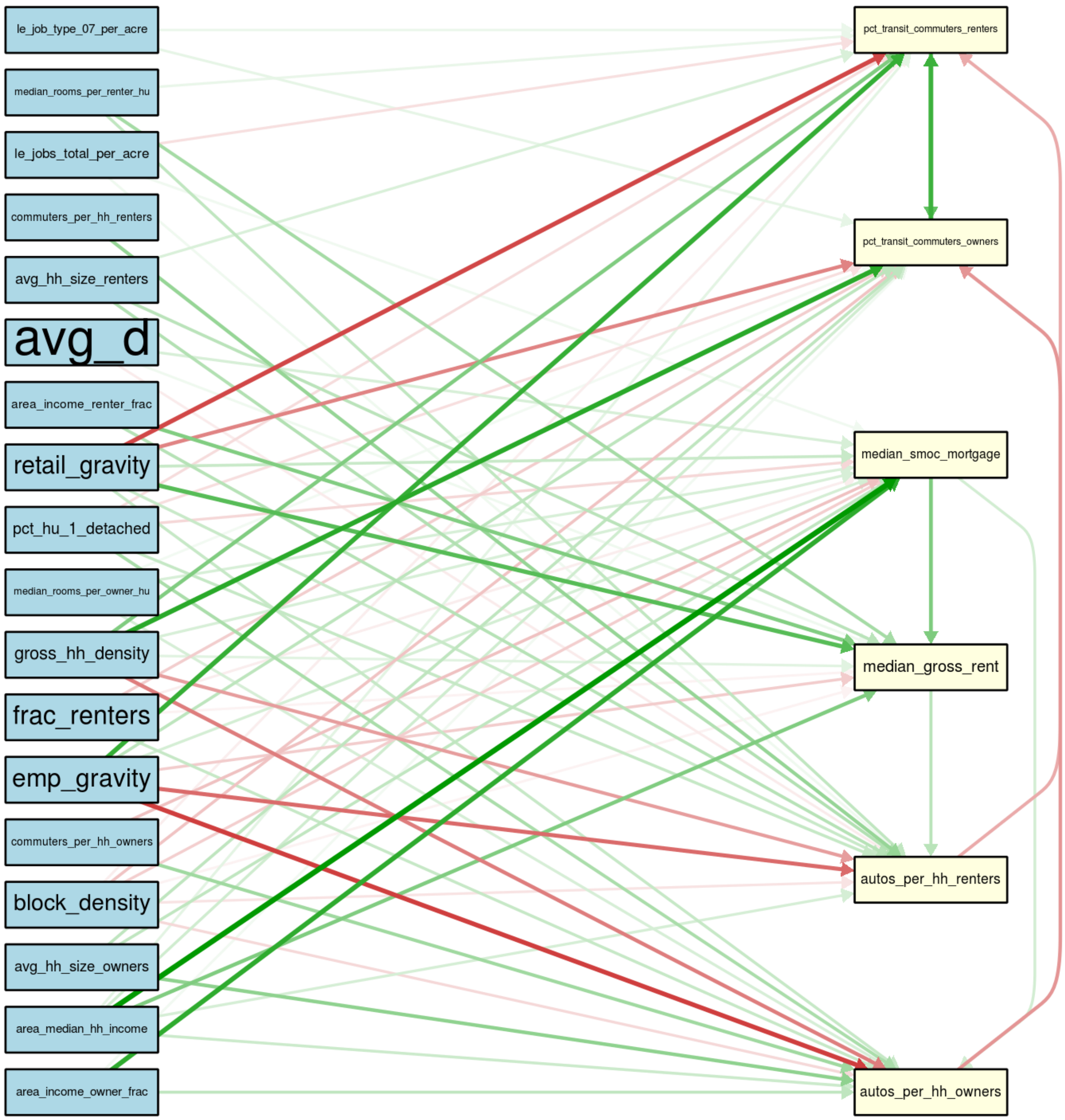
Appendix B: Path Diagrams

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

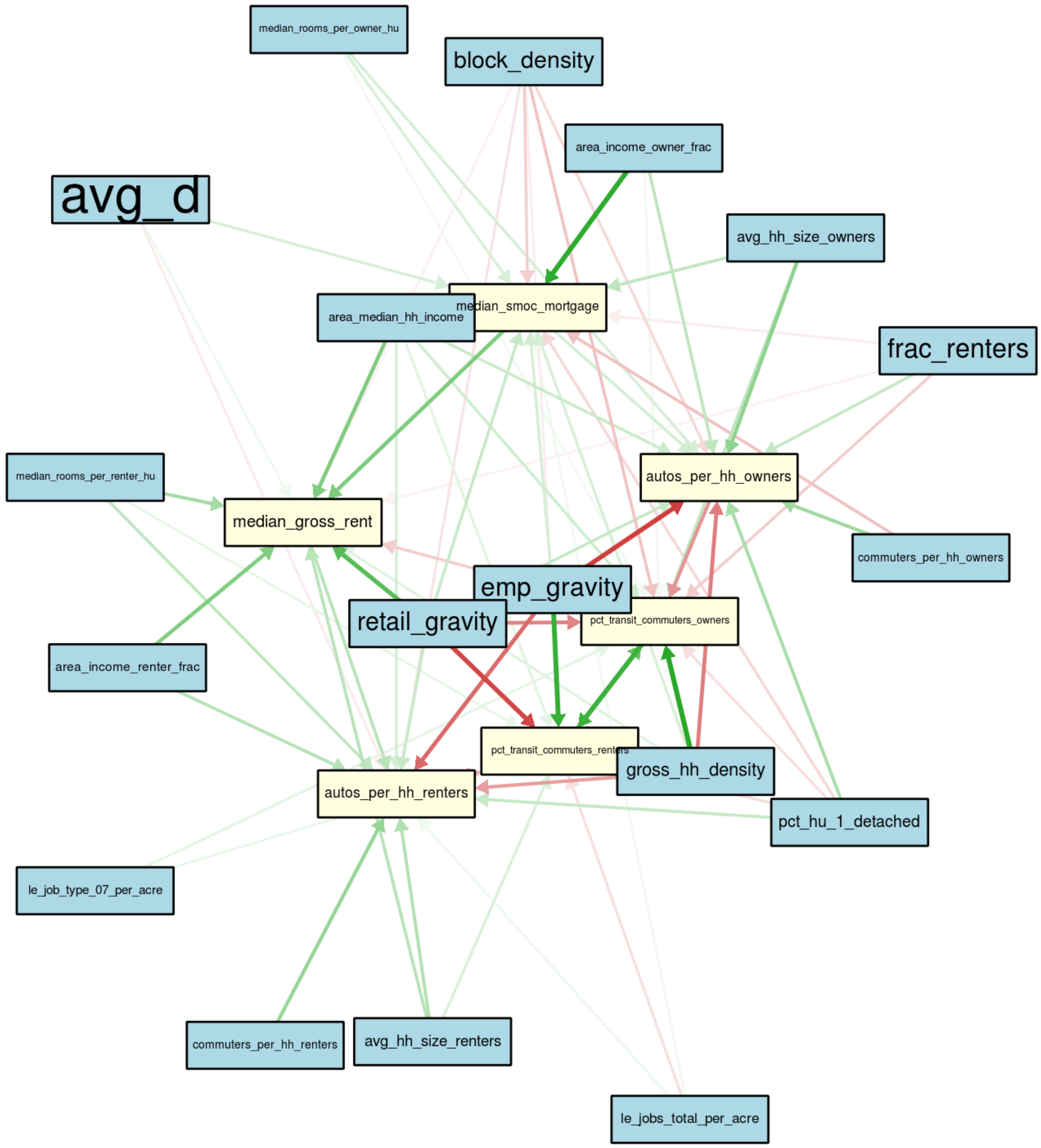
The width and darkness of the line indicates the strength of the relationship: wider darker lines indicate strong relationships while thinner lighter lines indicate weaker relationships. The path diagram illustrated in Figure 5 shows the values of the standardized variables used for LAIM Version 2 (Figure 6 is the same diagram but with a different layout).

Figure 5: Path Diagram for SEM Model



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

Figure 6: Path Diagram for SEM Model - Alternative Layout



Path Diagram Key

Line	Value
(dark green arrow)	0.40
(medium green arrow)	0.20
(light green arrow)	0.04
(light red arrow)	-0.04
(medium red arrow)	-0.20
(dark red arrow)	-0.40