H + T Index Methods

November 2022

H+T Index Methods

Table of Contents

Table of Tables

Table of Figures

Table of Equations

Introduction

The Center for Neighborhood Technology's Housing + Transportation (H+T®) Affordability Index (H+T Index) is an innovative tool that measures the true affordability of housing by calculating the transportation costs associated with a home's location. Planners, lenders, and most consumers traditionally consider housing affordable if the cost is 30 percent or less of household income. The H+T Index proposes expanding the definition of housing affordability to include transportation costs at a home's location to better reflect the true cost of households' location choices. Based on research in metro areas ranging from large cities with extensive transit to small metro areas with extremely limited transit options, CNT has found 15 percent of income to be an attainable goal for transportation affordability. By combining this 15 percent level with the 30 percent housing affordability standard, the H+T Index recommends a new view of affordability defined as combined housing and transportation costs consuming no more than 45 percent of household income.

The H+T Index was constructed using the measured housing cost and modeled transportation cost. The housing cost are obtained from the American Community Survey 5-year Estimate (2019 ACS) using the selected monthly ownership cost and the gross rent and combines each using the relative number of owner occupied households and renting households. The transportation model estimates three dependent variables (auto ownership, auto use, and transit use) as functions of 16 independent variables:

1. median household income,
2. average household size,
3. average commuters per household,
4. gross household density,
5. household intensity,
6. fraction of single family detached housing,
7. single family detached housing intensity,
8. fraction of rental housing units,
9. rental housing intensity,
10. employment intensity,
11. employment mix index,
12. block size,
13. bus transit connectivity index,
14. other (non-bus) transit connectivity index,
15. total available transit trips per week at peak times,
16. area of transit access shed, and
17. jobs within the transit access shed

To focus on the built environment's influence on transportation costs, the independent household variables (income, household size, and commuters per household) are set at fixed values to control for any variation they might cause. By establishing and running the model for a control household any variation observed in transportation costs is due to place and location, not household characteristics.

Transportation Cost Model Improvements

In our 2016 release made several improvements including the addition of rural block groups, an improvement of the transportation cost model, an improved method to derive household transit costs and interacting all variables that better capture the relationships between independent variables.

Those improvements are all included in the current iteration and a few improvements were made for this release as well; these include:

- Simplifying the transit connectivity index (TCI) and breaking it into bus and non-bus components,
- Adding peak hour transit frequency, intensity measures for the single family detached and rental housing units,
- Using a Quasibinomial fit for the percent of commuters using transit
- Using a logarithmic transformation of the dependent variable for the auto ownership and usage models

Geographic Level and Data Availability

The H+T Index was constructed at the Census block group level. Currently the H+T Index covers all 217,190 Census Block Groups in the US state and the District of Columbia. Due to incompatible and insufficient data Puerto Rico and other US territories were not included.

Data Sources

The H+T Index uses data from a combination of Federal sources and transit data compiled by the Center for Neighborhood Technology.

- 2015-2019 American Community Survey 5-year Estimate (2019 ACS) – an ongoing U.S. Census survey that generates data on housing characteristic, transportation use, community demographics, income, and employment.
- U.S. Census TIGER/Line Files – 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 and characteristic detail on age, earnings, industry distributions, and local workforce indicators. LODES data built on 2019 Census data are used here (Alaska, Arkansas, and Mississippi does not have 2014 LEHD data, 2016, 2018 and 2018 data were used respectively, assuming that the differences would be minimal).

- Average annual expenditures and characteristics of all consumer units, from the Consumer Expenditure Survey, 2006-2012 and 2019, used to inflate the cost of auto ownership from the 2010 data above.

- 2019 National Transit Database –fare box revenue and number of transit trips reported by agencies that receive federal assistance.
- AllTransit™ – a database of General Transit Feed Specification (GTFS) data developed by the Center for Neighborhood Technology, including bus, rail, and ferry service for both transit agencies that report their GTFS data publicly and those derived by CNT staff for agencies that do not.
- Odometer readings from The Illinois Department of Natural Resources - odometer data collected by Vehicle Emissions Testing Program.

Housing Costs

To calculate the H in the H+T Index, housing costs are derived from nationally available datasets. Median selected monthly owner costs for owners with a mortgage and median gross rent, both from the 2019 ACS, are averaged and weighted by the ratio of owner- to renter-occupied housing units from the tenure variable for every block group in a CBSA.

Transportation Cost Model

While housing costs are derived from 2019 ACS data, transportation costs, the T in the H+T Index, are modeled based on three components of transportation behavior—auto ownership, auto use, and transit use—which are combined to estimate the cost of transportation.

Basic Structure

The household transportation model is based on a multidimensional regression analysis, in which formulae describe the relationships between three dependent variables (auto ownership, auto use, and transit use) and independent household and local environment variables. Neighborhood level (Census block group) data on median household income, household size, commuters per household, household residential density, walkability and street connectivity, transit connectivity and access, and employment access and diversity were utilized as the independent or predictor variables.

To construct the regression equations, each predictor variable was tested separately; first to determine the distribution of the sample and second to test the strength of the relationship to the criterion variables. The regression analysis was conducted in a comprehensive way, ignoring the distinction between the local environment variables and the household variables to obtain the best fit possible from all the independent variables. The predicted result from each model was multiplied by the appropriate price for each unit—autos, miles, and transit trips—to obtain the cost of that component of transportation. Total transportation costs were calculated as the sum of the three cost components as follows:

$$\text{Household T Costs} = [C_{AO} * F_{AO}(\text{X})] + [C_{AU} * F_{AU}(\text{X})] + [C_{TU} * F_{TU}(\text{X})]$$

Equation 1: Cost of Transportation

Where:

C = cost factor for each of the transportation components (i.e., dollars per mile)
F = modeled dependent variables as a function of the independent variables (F_{AO} is auto ownership, F_{AU} is auto use, and F_{TU} is transit use)

Dependent Variables

Auto Ownership

For the dependent variable auto ownership, the regression analysis was fit using measured data on auto ownership obtained from the 2019 ACS. Aggregate number of vehicles available by tenure defined the total number of vehicles, and tenure defined the universe of occupied housing units. Average vehicles per occupied housing unit were calculated at the block group level. A logarithmic transformation was used to better represent the auto ownership, since it cannot have a negative value.

Auto Use

For the dependent variable auto use, the regression analysis was fit using measured data on the amount households drive, vehicle miles traveled (VMT) per automobile. Odometer readings from 2018 through 2020 odometer readings were acquired in Illinois for the Chicago and St. Louis metro areas. Data were matched for over 660,000 records (two records for each individual vehicle identification number (VIN)) and the change provided VMT estimates. The dataset represents a diverse set of place types from rural areas to large cities and provides a very good data set to calibrate the model. Data obtained were geographically identified with ZIP+4™ and then assigned to Census block groups.

Because of the Covid lock down driving in 2020 was reduced, the required a complicated formula to back out the driving that happened in 2019 since. See upcoming blog on how this calculation was done.

The final value of VMT includes an additional factor of eight percent to compensate for the fact that the vehicles in this sample were all five years old or older. This factor is obtained from the research commissioned and published by US HUD and US DOT to develop the Location Affordability Index.[1]

Like to the auto ownership model the VMT per household a logarithmic transformation was used to better represent the household driving, since it cannot have a negative value and is skewed right.

Transit Use

Because no direct measure of transit use was available at the block group level, a proxy was utilized for the measured data representing the dependent variable of transit use. From the 2019 ACS, means of

[1] See http://www.locationaffordability.info/LAPMethodsV2.pdf page 24.

transportation to work was used to calculate a percent of commuters utilizing public transit. Previously a ordinary least square regression was used to fit these data, however, that is not appropriate for a distribution of probability, which this variable essentially is. Thus, a more appropriate approach was employed using a Quasibinomial fit.

Independent Variables

The independent variables used in all three models are divided into those that describe the households living in each Census Block Groups, and those that describe its environment. Table 1, below, summarizes these inputs, which are further described in the next sections.

Table 1: Independent Variables Overview

Variable	Description	Data Source	Type
Median HH Income	Median household Income in the block group	2019 ACS	Household
Commuters/HH	Workers per household who do not work at home	2019 ACS	Household
Avg. HH Size	Average number of people per household	2019 ACS	Household
Gross Household Density	Number of households divided by the land area in the Census block group	2019 ACS, TIGER/Line files	Neighborhood Characteristic (housing density)
Household Intensity	Households summed divided by the distance squared in miles between block group by (the households in the block group are not included)	2019 ACS, TIGER/Line files	Neighborhood Characteristic (housing density)
Fraction of Rental Housing units	Fraction of occupied housing units with rental tenure	2019 ACS,	Neighborhood Characteristic (housing density)
Fraction of Rental Housing UNITS' Intensity	Fraction of occupied housing units with rental tenure divided by the distance squared in miles between block group by (the households in the block group are not included)	2019 ACS, TIGER/Line files	Neighborhood Characteristic (housing density)
Fraction of Single-Family Detached Housing	Fraction of Single-Family detached housing units in the block group	2015 AC	Neighborhood Characteristic (housing density)
Fraction of SINGLE-FAMILY Detached Housing Intensity	Fraction of Single-Family detached housing units divided by the distance squared in miles between block group by (the households in the block group are not included)	2019 ACS, TIGER/Line files	Neighborhood Characteristic (housing density)
Employment Intensity	Jobs summed by blocks divided by the distance squared in miles (if less than one mile not scaled)	Census LEHD-LODES	Neighborhood Characteristic (employment)
Employment Mix Index	Sum of 20 different employment types each scaled by a coefficient that are optimized using auto ownership	Census LEHD-LODES	Neighborhood Characteristic (employment)
Block Size	Average number acre per block	TIGER/Line files	Neighborhood Characteristic (walkability)
Bus Transit Connectivity Index	Sum of buses per week scaled by overlap of 3/8 mile buffer around routs that intersects the block group	CNT AllTransit	Neighborhood Characteristic (transit)

Other Transit Connectivity Index	Sum of rail, ferry, and other modes (not bus) trips per week scaled by overlap of every 1 mile buffer around routs that intersects the block group	CNT AllTransit	Neighborhood Characteristic (transit)
TAS Area	Areas in Acres riders in the block group can access in 30 minutes with 1 or no transfers for all the transit stations within a ¼ mile of the block group	CNT AllTransit	Neighborhood Characteristic (transit)
TAS Jobs	The total number of jobs in TAS	CNT AllTransit	Neighborhood Characteristic (transit/employment)
Average Available Transit Trips per Week at Peak Times	Number of possible transit rides within the block group and a ¼ mile of its border that occur at peak time of day.	CNT AllTransit	Neighborhood Characteristic (transit)

Independent Variables: Household Characteristics

The 2019 ACS, at the block group level, serve as the primary data source for the independent variables pertaining to household characteristics.

Median Household Income:

Median household income is obtained directly from the 2019 ACS.

Average Household Size:

Average household size was calculated using total population in occupied housing units by tenure.

Average Commuters per Household:

Average commuters per household was calculated using the total workers 16 years and over who do not work at home from means of transportation to work and tenure to define 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 occupied housing units to total population was used to scale the count of commuters to better represent those living in households.

Independent Variables: Neighborhood Characteristics

Household Residential Density

In previous versions of the H+T Index household density was found to be one of the most significant variables in explaining the variation in auto use, auto ownership, and transit use. Various definitions of density have been constructed and tested, but net residential density (households per residential acre) was the primary metric used. No national data source of detailed land use data exists so previous versions of the household transportation cost model defined residential density as the average number of households per residential acre for the Census blocks within the block group weighted by count of households. Total households obtained at the block level from the 2010 US Census and TIGER/Line files were used to define blocks. However, since this iteration is using data from the 2019 ACS, the 2010 data

is not compatible. Thus, several metrics were developed to estimate how household transportation behavior is driven by household density and concentration.

Gross Household Density

Gross household density is calculated from the 2019 ACS. It is simply the number of households in a census block group divided by the area of land within the block group

Household Intensity

The Household Intensity is constructed using a gravity model which considers both the quantity of, and distance to, all households, relative to any given block group. Using an inverse-square law, intensity is calculated by summing the total number of households divided by the square of the distance to those households but does not include the households within the block group. This quantity allows us to examine the intensity of housing development in the region around the block group.

The Household Intensity is calculated as:

$$H \equiv \sum_{i=1}^{n} \frac{hh_i}{r_i^2}$$

Equation 2: Household Intensity Definition

Where:

H is the Household Intensity for a given Census block group
n is the total number of Census blocks (not including the given Census block group)
hh_i is the number of households in the i^{th} Census block
r_i is the distance (in miles) from the center of the given Census block group to the center of the i^{th} Census block

As households get farther away from the Census block group their contribution to the Household Intensity is reduced; for example, one household in a Census block group a mile away adds one, but a household 10 miles away adds 0.01. All households in all US Census blocks groups are included in this measure. However, to expedite the calculation, the calculation uses the[2]:

- State totals when the state is not the same as the given block group and is more than 1256 miles away,
- County totals when the county is not the same as the given block group and is more than 165 miles away, and
- Census tract totals when the tract is not the same at the given block group and is more than 34 miles away.

Fraction of Rental House Housing Units

The fraction of occupied housing units that are rental rather than owned by the occupant is calculated using the 2019 ACS data by dividing the number of occupied housing units with tenure defined as rental by the total number of occupied housing units in the Census block group.

[2] These distance thresholds were developed using the average distance between the geographic entities time five.

Fraction of Single Family Detached Households

The fraction of single family detached households is calculated using the 2019 ACS data by dividing the number of households living in single family detached housing by the total number of households in the Census block group.

Street Connectivity and Walkability

Measures of street connectivity have been found to be good proxies for pedestrian friendliness and walkability. Greater connectivity created from numerous streets and intersections creates smaller blocks and tends to lead to more frequent walking and biking trips, as well as shorter average trips. Three measures of street connectivity — block size, intersection density, and block perimeter — have been found to be important drivers of household travel behavior. However, these three measures are so interrelated only block size was included. The resulting models have essentially equivalent R^2 values compared to when the other measures are included and thus have comparable goodness of fit.

Block size

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

Employment Access and Diversity

Employment numbers are calculated using Longitudinal Employer-Household Dynamics (LEHD) Origin Destination Employment Statistics (LODES) at the Census block group level. The Longitudinal Employer-Household Dynamics (LEHD) program is part of the Center for Economic Studies at the U.S. Census Bureau.

Employment Intensity

The Employment Intensity is constructed using a gravity model that factors in the quantity of, and distance to, all employment destinations, in relation to any given block group. Using an inverse-square law, the Employment Intensity is calculated by summing the total number of jobs divided by the square of the distance to those jobs. This method provides more information than a simple job density measure, in that it includes the accessibility to jobs outside a given Census block group. In addition to measuring access to jobs, it also provides a measure of economic activity created by those jobs.

The Employment Intensity is calculated as:

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

Equation 3: Employment Intensity Definition

Where:

E is the Employment Access for a given Census block group
n is the total number of Census blocks
p_i is the number of jobs in the i^{th} Census block
r_i is the distance (in miles) from the center of the given Census block group to the center of the i^{th} Census block

The proximity of jobs to the Census block group determines their contributive value to the Employment Intensity. For example, one job a mile away adds one, but a job 10 miles away adds 0.01. The measure includes all jobs in all US Census blocks. The index employs the following parameters to accelerate the calculation:[3]

- State totals when the state is not the same as the given block group and is more than 1,256 miles away,
- County totals when the county is not the same as the given block group and is more than 165 miles away, and
- Census tract totals when the tract is not the same at the given block group and is more than 34 miles away.

Employment Mix Index

The model includes an Employment Mix Index which measures employment diversity in addition to total number of jobs. It is produced by taking the weighted sum of the gravity measure of each type of job (out of 20 total types). The benefit of looking at the mix of employment options can be seen in the R^2 value for the transit use model. The transit use model when transit data is not available produces an R^2 value of 78.8%, but when the employment mix index is included the R^2 increases to 80.7%.

Table 2 lists the 20 employment categories derived from the 2019 LEHD, the variable is the LEHD variable name, the function is the linearization function used in the fit, and what the trend is for auto ownership. The variable is the gravity measure of the jobs of the given type (see above). The weight is determined by regressing all the other independent variables and these 20 against autos per household. The table is ordered in decreasing order by how much each variable changed the R^2 of the fit, note that the individual R^2 is the goodness of fit value if this is the only variable used to fit auto ownership and the incremental R2 is the R2 of the Auto ownership fit when the variable is added to the fit. The household variable were included in the fit to control for that variation, but not shown in the table since they are not included in the employment mix index. The *Trend* column indicates what happens to the value of the Employment Mix Index when the fraction of the given employment type increases.

[3] These distance thresholds were developed using the average distance between the geographic entities and factor determined such that the calculation remains consistent with using the block groups for a small representative sample.

Table 2: Summary of Employment Type and Weighting for Employment Mix Index

NAICS Code(s)	Variable	Function	Weight	Trend	Individual R^2	Incremental R^2
53-Real Estate and Rental and Leasing	CNS11	√ x	-.0078	Reduced	31.5%	31.5%
81-Other Services [except Public Administration]	CNS19	√ x	-.00654	Reduced	31.5%	67.9%
56-Administrative and Support and Waste Management and Remediation Services	CNS14	√ x	.00351	Increased	24.9%	69.2%
62-Health Care and Social Assistance)	CNS16	ln(1+x)	-.0532	Reduced	23.1%	70.8%
61-Educational Services	CNS15	x^2	-1.20x10-10	Reduced	20.1%	71.4%
31-33-Manufacturing	CNS05	1/x	-1.86	Increased	0%	72%
11-Agriculture, Forestry, Fishing and Hunting	CNS01	ln(x)	.0192	Increased	0.1%	72.3%
48-49-Transportation and Warehousing	CNS08	x^2	-1.54 x10-10	Reduced	5.8%	72.6%
21-Mining, Quarrying, and Oil and Gas Extraction	CNS02	1/(1+x)	-.133	Increased	1.1%	72.8%
42-Wholesale Trade	CNS06	√ x	.00229	Increased	20.1%	72.9%
72-Accommodation and Food Services	CNS18	√ x	.00161	Increased	28.7%	73%
52-Finance and Insurance	CNS10	√ x	-.00112	Reduced	29.4%	73%
44-45-Retail Trade	CNS07	√ x	-.00087	Reduced	26.5%	73.1%
22-Utilities	CNS03	1/(1+x)	.15	Reduced	5.4%	73.1%
23-Construction	CNS04	√ x	-.00082	Reduced	20.8%	73.1%
54-Professional, Scientific, and Technical Services	CNS12	√ x	.00047	Increased	25.8%	73.2%
71-Arts, Entertainment, and Recreation	CNS17	1/(1+x)	.19	Increased	3%	73.2%
92-Public Administration	CNS20	1/x	-.23	Reduced	0.6%	73.2%
51-Information	CNS09	1/(1+x)	.18		3.6%	73.2%
55-Management of Companies and Enterprises	CNS13	1/(1+x)	-.03	Reduced	2.7%	73.2%

The calculation for the raw employment mix is:

$$R \equiv \sum_{i=1}^{13} W_i \times F_t(e_i)$$

Equation 4: Definition of Raw Employment Mix

Where:

R is the Raw Employment Mix for a given Census block group

i is the employment category

w_i is the weight for the i^{th} employment category

F_i is the linear transformation function for the i^{th} employment category (ln(x) for all except 1/x for NAICS coded 55 and 71).

e_i is the value of the variable in Table 2 for the i^{th} employment category

The full calculation is then evaluated using the following formula.

$$I_{Emix} \equiv 100 \times \frac{R - R_{min}}{R_{max} - R_{min}}$$

Equation 5: Definition of Employment Mix Index

Where:

I_{Emix} is the Employment Mix Index for a given Census block group
R is the Raw Employment Mix for a given Census block group
R_{min} is the minimum value of the Raw Employment Mix for all Census block groups
R_{max} is the maximum value of the Raw Employment Mix for all Census block groups

This index is calculated of all Census block groups in the country as a number from zero to 100; the Figure 1 shows the distribution of values for this index for only the Census block groups in the sample used the H+T index, as does Figure 2: Histogram of Employment Mix Index Where Index is Less Than FortyFigure 2 but this shows only the lower values illustrating the long tail. This comes from very remote locations such as in the Aleutian Islands in Alaska, and the outer islands in Hawai'i.

Figure 1: Histogram of Employment Mix Index

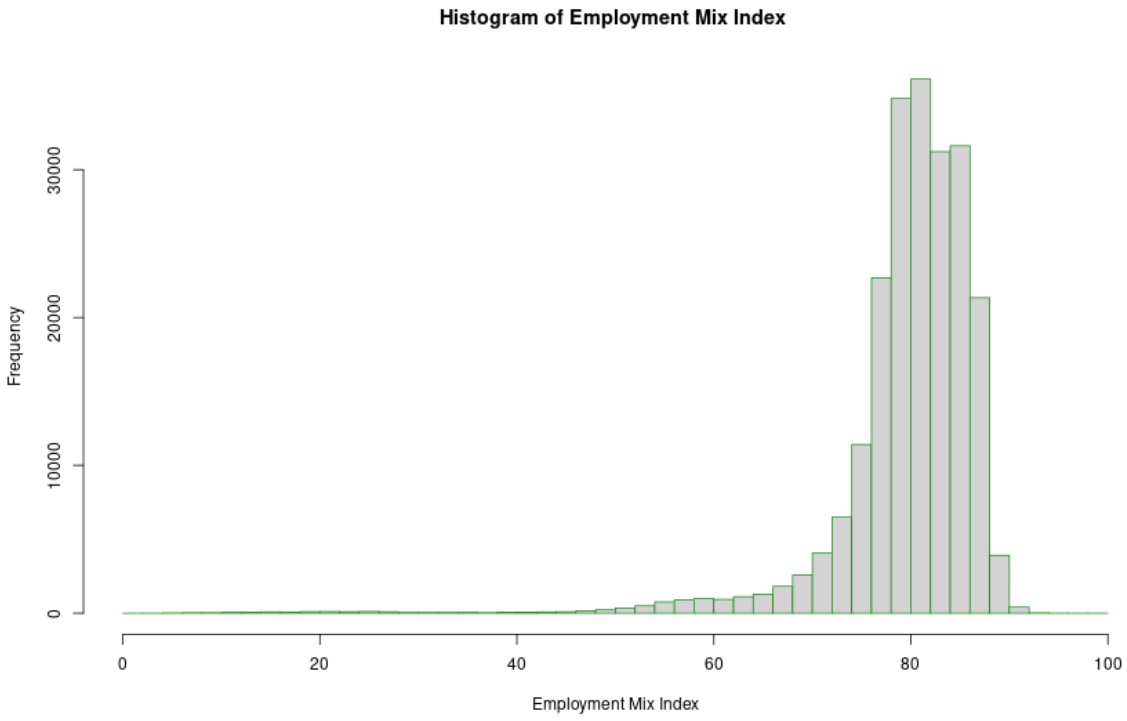
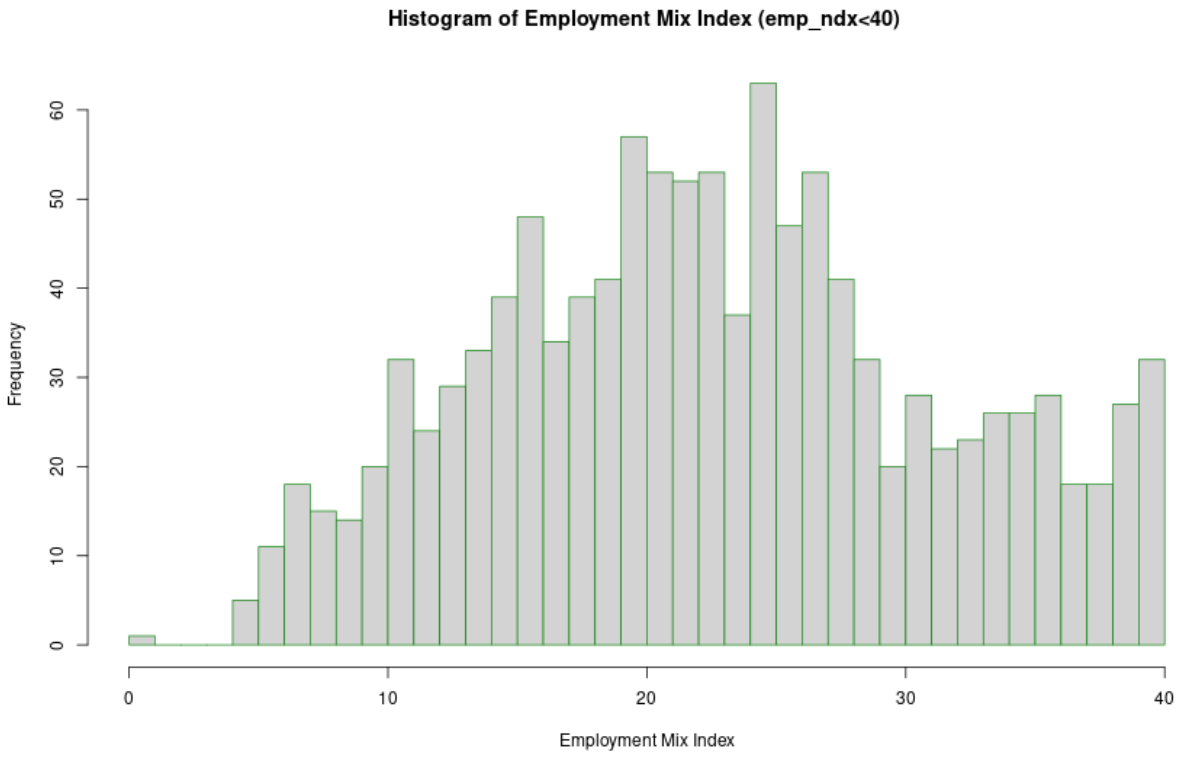


Figure 2: Histogram of Employment Mix Index Where Index is Less Than Forty



Transit Access and Connectivity

Transit access is measured through *General Transit Feed Specification* (GTFS) data collected and synthesized by CNT. In addition to the publicly available GTFS data (provided by many, but not all, transit agencies) CNT has created GTFS structured datasets utilizing online transit maps and schedules. In many cases, CNT has directly contacted transit agencies to obtain more specific information on stop locations and schedules. All GTFS data is merged into a dataset through a CNT tool known as AllTransit™ Data Builder. AllTransit is an online tool that facilitates the collection, normalization, aggregation, and analysis of GTFS data to determine fixed-route transit service.

To date, CNT has compiled stop, station, and frequency data for bus, rail, and ferry service for all major transit agencies in regions with populations greater than 100,000. Attachment A lists the transit agencies for which data has been compiled. In regions where data is not available, CNT has assumed that the transit service is not large enough to affect the fits, thus zero is assumed for all transit inputs.

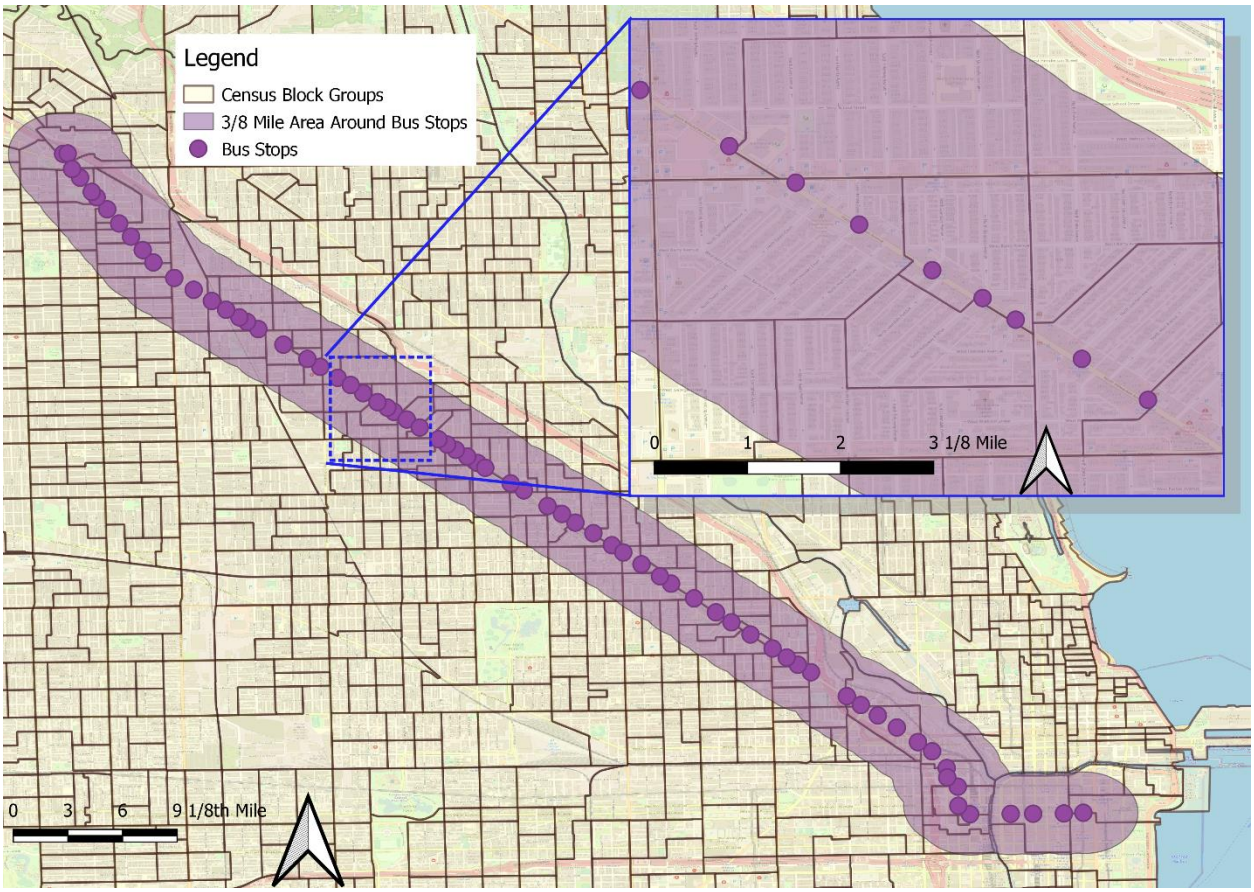
Five measures of transit access are used in the model: two Transit Connectivity Indexes (TCI) one for bus and one for non-bus or other, Transit Access Shed Area (TAS area), Transit Access Shed Jobs (TAS Jobs), and Average Available Transit Trips per Week as peak time. The TCI estimates how many transit opportunities are within walking distance of a census block group per week. The TAS is a proxy measure for how far one can travel in 30 minutes on transit (TAS), with one or no transfers, the area is the total

area of that best-case-trip using ¼ mile circle about each stop and the jobs are the sum of the total number of jobs within the TAS.

Transit Connectivity Indexes

The Transit Connectivity Index is a measure of access to bus stops or rail stations that CNT developed specifically for use in the household transportation cost model. In previous iterations the TCI used 1/8-mile rings around stations/stops and developed a weighted sum of these rings, this however, was a complicated and more nuanced method but it did not add that much more to the overall fits, therefore, the TCI has been simplified using single fixed area around the stations/stops. To calculate this measure two different areas were developed, one for bus transit lines, and one for all other types[4] of transit. For bus routes a three-eighths-mile in width was used (Figure 3 shows below an example for a single route in Chicago) and for the other types a one-mile buffer was employed.

Figure 3: Illustration of 3/8th mile area around Bus Route (For a Single Route in a Chicago)



The following formula is used for each transit route that intersects in to obtain the bus TCI for each block group.

[4] For other we include, light rail, heavy rail, commuter rail, ferry terminals, and other stations such as vernaculars and cable cars.

$$Bus/Other\ TCI = \sum_{i=1}^{n} \frac{L_{i,}F_{i,}}{B}$$

Equation 6: Bus/Other TCI Calculation

Where:

L is the land area of the block group covered by access zone
F is the service frequency value (trips/week)
B is the total block group land area
n is the total intersecting transit routes access zones.

These values are calculated for every block group that a given zone intersects; meaning that in well-served block groups there will be values for zones corresponding to multiple stops, and for Block Groups that have no adjacent transit the value is zero. The following figures show the distribution of the two types of TCIs excluding the Block Groups that have zero value.

Figure 4: Histogram of Bus TCI for Block Groups that Intersect at Least One Bus Service Area

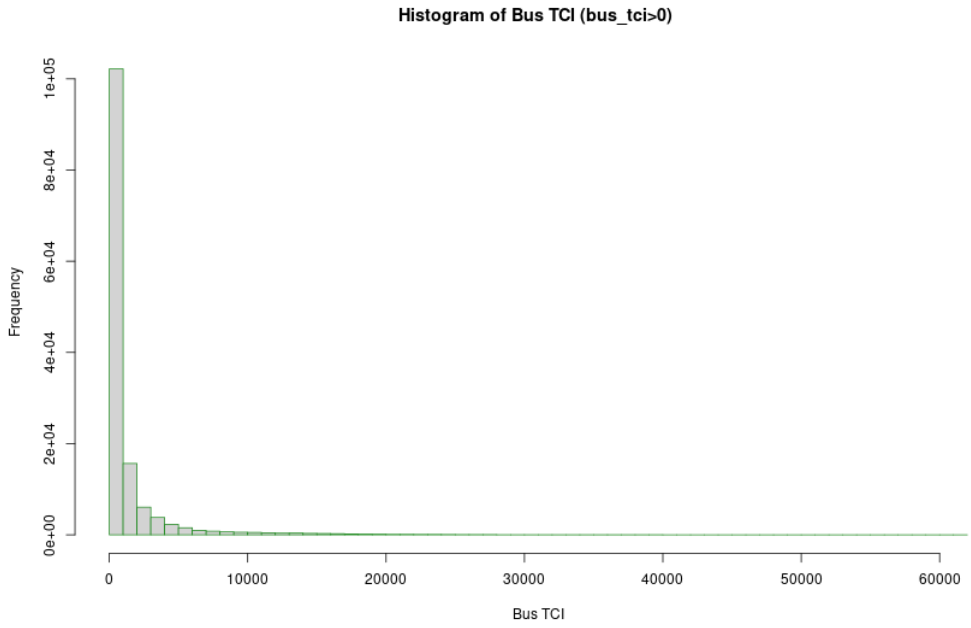


Figure 5: Histogram of Other TCI for Block Groups that Intersect at Least One Other Service Area

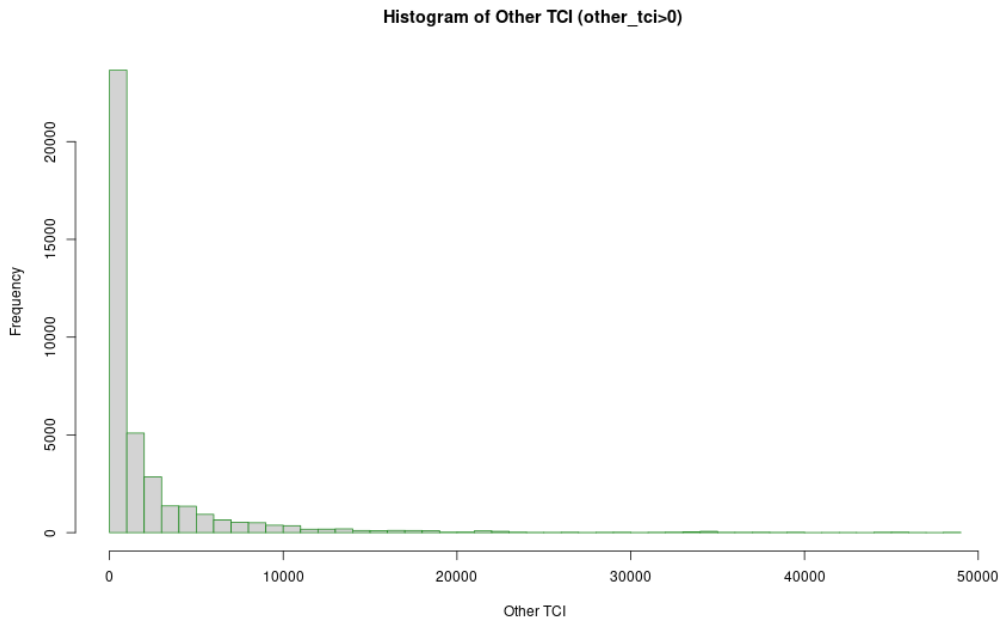


Figure 6: Histogram of Bus TCI for Block Groups that Intersect at Least One Bus Service Area and that have value greater than 10,000 trips/week

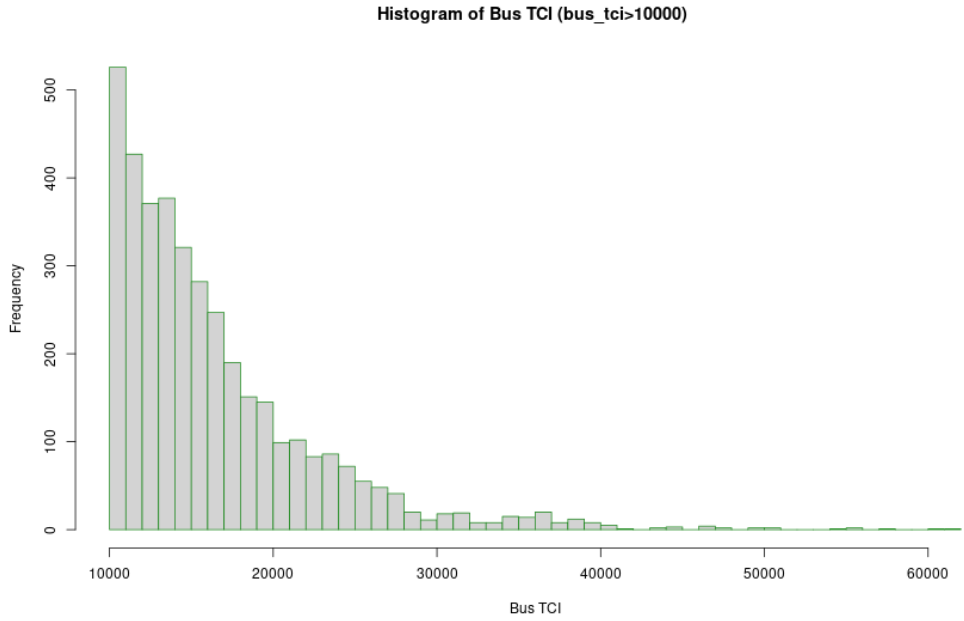
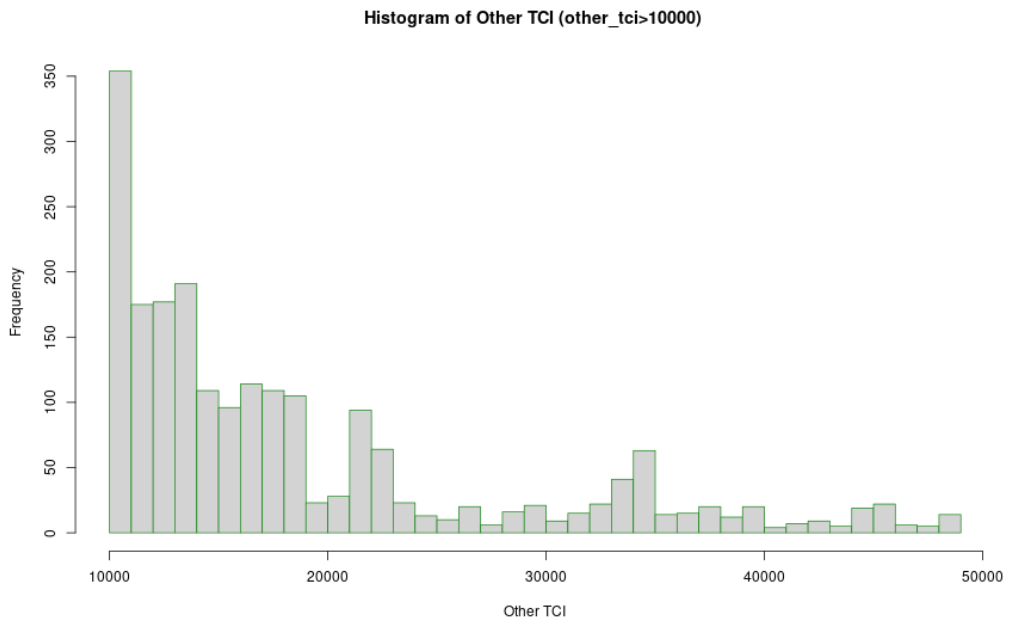


Figure 7: Histogram of Other TCI for Block Groups that Intersect at Least One Other Service Area and that have value greater than 10,000 trips/week



The farther an access zone is from its transit node, the less of a contribution it makes to the level of access in any block group it intersects. Thus, the TCI value is weighted sum of all the transit available near the block group scaled by the amount of area covered by fraction of area covered by given distance to transit stops/stations. This measure captures the station/stop availability as well as the frequency of service.

Transit Access Shed Area

The Transit Access Shed (TAS) is defined as a geographic area accessible within 30 minutes by public transportation. This measure was derived from the AllTransit GTFS data. For each transit stop, all stops that can be reached within 30 minutes were identified. One transfer within a quarter mile of a stop was allowed, and all transfers were padded with 10 minutes of walking and/or waiting. The stops reachable within 30 minutes were based on the minimum travel time between the two stops, allowing the inclusion of more distant stops that are reachable within 30 minutes via express service. For each origination stop, a quarter-mile buffer was created around the destination stops. Based on the location of the originating stop, the access shed was then aggregated for each stop to the block group by including stops that were within the block group or within a quarter of a mile of its boundary. Finally, the accessible area or Transit Access Shed is calculated by summing the areas of the quarter-mile buffers around every stop that is within 30 minutes as defined above. To assign a value to a Census block group, the Transit Access Shed for all stops within walking distance of the block group are merged into one grand shed. This area is then assigned as the block group's Transit Access Shed.

Transit Access Shed Jobs

Transit Access Shed Jobs is the total number of jobs within the TAS. The count of jobs was obtained from the Census LEHD-LODES data.

Average Available Peak Transit Trips per Week

Average Available Transit Trips per Week is the average frequency of service from the AllTransit GTFS data, for all stops within the Census block group or within a half mile of it borders between the hours of 7:00am and 9:00am.

Household Transportation Regression Analysis

As in the previous versions of this index, in this iteration this non-linearity of the independent variables is compensated for by using simple transformation functions. These functions (Linear (x), Square Root ($\sqrt{x}$), Natural Log ($\ln(x)$), Square (x^2) and Inverse ($1/x$)[5]) are used to give the best fit using an ordinary least square fit. The final fit uses eleven independent variables and is broken up into six independent models. For the percent of commuters using transit for their journey to work, a Quasibinomial regression is use, and for the autos per household and household VMT models the dependent variable uses a logarithmic transformation.

An ordinary least square regression including all interaction terms (that are statistically significant) was used to estimate the fit coefficients; the equation that will estimate the dependent variable from independent variables is:

$$D = I + \sum_{i=1}^{n} C_i \times f_i(x_i) + \sum_{i=1}^{n} \sum_{j=i}^{n} C_{ij} \times f_i(x_i) \times f_j(x_j)$$

Equation 7: Equation for Estimating Dependent Variable from Coefficients and Independent Variables

[5] For the natural log and the inverse functions a "safe" version of the function was used where the value of x can be zero – that take the form of ln(1+x) and 1/(1+x).

Where:

D is the linearized dependent variable for a given Census Block Group i.e. Autos per Household
I is the Intercept – obtained in the regression
i is the index of the independent variable i.e. i goes from 1 to 10 for a regression that had 10 independent variables
j is the index of the independent variable i.e. j goes from i to 10 for a regression that had 10 independent variables and so as to include all interaction terms only once
C_i is the fit coefficient for the i^{th} independent variable
C_{ij} is the fit coefficient for the i^{th} j^{th} interaction variable
f_i is the linearization transformation function
x_i is the value of the i^{th} independent variable

Note that, because of the use of the quasibinomial fit for the percent of commuters using percent commuter using transit for their journey to work to work the final is calculated uses the function:

$$R = \frac{e^D}{(1 + e^D)}$$

And for the auto ownership and household VMT the final values are the exponential of D, since the logarithmic transformation was used in the fit.

Choosing Linear Transformation Functions for the Independent Variables

To address the nonlinear nature of the relationships between the independent and the dependent variables, a linear transformation function was chosen. To choose the best transformation each variable was tested to determine which transformation resulted in the best fit. The example below considers how median household income drives auto ownership. Figure 8 and Figure 9 average household size is nonlinear. However, the relationship between auto ownership and the inverse (1/x) shows a more linear relationship; note the increase in R^2 by over four percentage points. This technique was then repeated for every variable for every model to select the optimal transformation.

Figure 8: Autos/HH vs. Avg. Household Size

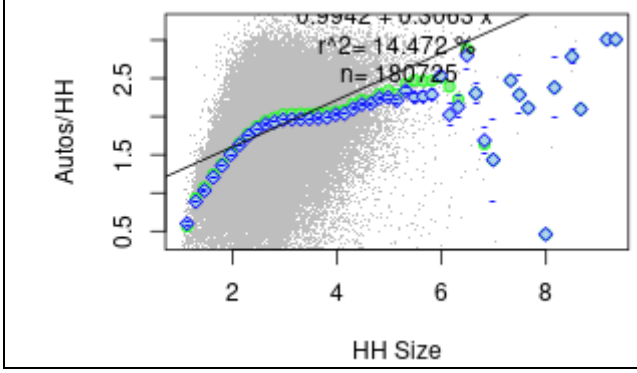


Figure 9: Autos/HH vs. the inverse of Avg. Household Size

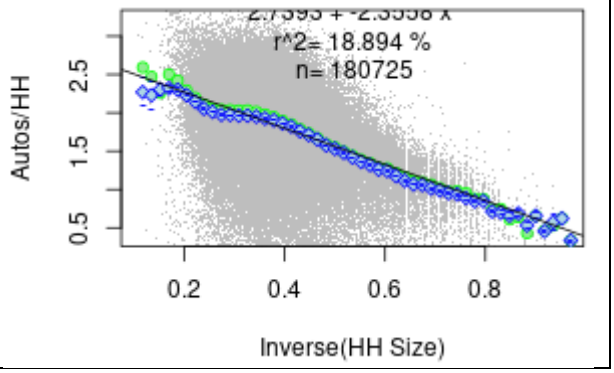


Table 3: Summary of Linearization Functions Used in the Three Transportation Regressions

Variable	Auto/HH F(x)	HH VMT F(x)	% Transit Commute F(x)
Median HH Income	1/(1+x)	1/(1+x)	1/x
Commuters/HH	x	1/(1+x)	√ x
Avg. HH Size	1/x	1/(1+x)	x
Gross Household Density	x	√ x	√ x
Household Intensity	x^2	√ x	√ x
Fraction of Rental Housing units	x	x	x
Fraction of Rental Housing Unit Intensity	LN(1+x)	LN(1+x)	x
Fraction of Single-Family Detached Housing	√ x	√ x	1/(1+x)
Fraction of Single-Family Detached Housing Intensity	x^2	LN(1+x)	x^2
Employment Intensity	x	LN(x)	√ x
Employment Mix Index	x^2	x^2	√ x
Block Size	LN(1+x)	1/(1+1/x)[6]	LN(1+x)
Bus Transit Connectivity Index	x	√ x	√ x
Other Transit Connectivity Index	x	√ x	LN(1+x)
TAS Area	√ x	LN(1+x)	x
TAS Jobs	x	LN(1+x)	x
Average Available Transit Trips per Week at Peak Times	√ x	LN(1+x)	LN(1+x)

Choosing Independent Variables

The method to test the statistical significance of variables and to determine those that would reduce the multicollinearity of the set of independent variables, and all the possible interaction terms is described below:

- Initially the fit is done including all possible variable combinations. Then an iterative process is conducted where the least statistically significant variable or combination of variables is dropped, examining the probability that the value is consistent with zero (Pr(>|t|)).
- Then the fit is redone, and the least statistically significant variable or combination of variables is dropped, this continues until the least statistically significant variable or combination of variables' probability that it is consistent with zero is less than 5%.
- This is done in an iterative manner because simply deleting all the variable or combination of variables that have Pr(>|t|))>5% would eliminate more variables than necessary, since all this set of variables and combination of variables display a very high level of multicollinearity.
- Once this process is completed, another is applied that adds each eliminated variable back into the regression equation to see if by eliminating all other variables it may become statistically significant again.

Transportation Cost Calculation

The transportation model in the H+T Index estimates three components of travel behavior: auto ownership, auto use, and transit use. To calculate total transportation costs, each of these modeled

[6] Household VMT model uses block density (1/block size) rather than block size.

outputs is multiplied by a cost per unit (e.g., cost per mile) and then summed to provide average values for each block group.

Auto Ownership and Auto Use Costs

Auto ownership and use costs are derived from research conducted by HUD and DOT using the Consumer Expenditure Survey (CES) from the US Bureau of Labor Statistics. The research is based on the 2005-2010 waves of the CES, and costs are estimated for autos up to ten years old. Because expenditures are represented in inflation-adjusted 2010 dollars using the Consumer Price Index for all Urban Consumers (CPI-U), an inflation factor is applied to estimate the cost of auto ownership into 2015 dollars. The factor used is derived from the CES; the average expenditure in 2010 is $2,588 and in 2019 it is $4,393, thus the factor applied is 1.697.

Expenses are then 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.

GHG Estimates

Given the estimate of household VMT, the greenhouse gas (GHG) emissions can be estimated. Using the fleet average fuel efficiency of 21.6 miles/gallon[7], and the emission factor of 8.92 kg/gallon of gasoline, and a factor of 1.015 of CO_2 to CO_2e for auto use[8] the average household CO_2e (GHG_{hh}) is calculated using the estimated VMT_{bg} in a block group (in this case we use the median income, average household size and average households commuters in the block group for the control household rather than the "typical household[9]") by the following formula:

$$GHG_{hh} = \frac{VMT \times 8.92kg\ CO_2/gallon \times 1.015\ CO_2e/CO_2}{21.6mpg}$$

Equation 8: Equation for Estimating Greenhouse Gas Emission per Household

This is then used to calculate the total the metric displayed as "Annual GHG per Household" and when multiplied by the total number of household and divided by the land area in a Census Block group as the metric "Annual GHG per Acre."

[7] See Transportation Energy Data Book: Edition 35 (http://cta.ornl.gov/data/tedb35/Edition35_Chapter04.pdf) bottom of table 4.3 – there is a new edition to this (https://tedb.ornl.gov/wp-content/uploads/2022/03/TEDB_Ed_40.pdf#page=97) that states that the MPG is 23.5 for 2019, however, the old number 21.6 was used in these calculations.
[8] See http://www.epa.gov/cleanenergy/energy-resources/refs.html
[9] See typical household definition in the Constructing the H+T Index section below

Transit Use Costs

The 2015 National Transit Database (NTD) served as the source for transit cost data. Specifically, directly operated and purchased transportation revenue were used. The transit revenue, as reported by each of the transit agencies in the 2015 NTD, was assigned to agencies and related geographies where GTFS data were collected. This transit revenue was allocated to the counties served based on the percentage of each transit agency's bus and rail stations weighted by the number of trips provided within each county served. For example, if a transit agency had a total of 500 bus stops and 425 of those stops were in county A, and 75 stops extend into a neighboring county B, and all stops are served at the same level of frequency, county A received 85 percent of the transit revenue and county B received 15 percent.

To estimate average household transit costs, the modeled percentage of transit commuters and total households in each block group was used. Each county's estimated transit revenue was assigned to block groups on this basis. The block group number of transit commuters is calculated and summed to estimate the total number of transit commuters in the county. The county-wide transit revenue is then allocated to block groups based on the proportion of the county's commuters living there. The average household transit cost for each block group is calculated by dividing the block group's allocation of transit revenue by number of households.

This same method was used to estimate the average number of household transit trips for each block group. Using the total unlinked trips from the 2015 NTD, this measure was estimated using allocation the total number of annual trips in each metropolitan area proportionally to block groups based on number of households and the percent of journey to work trips.

There are several counties for which GTFS data are not available and/or there was no revenue listed in the 2015 NTD. In these cases, the national averages from previous paragraphs were used for these counties. The average transit costs and trips were then allocated to the block group level based on the percentage of transit commutes and household commuter counts. The result was an average household transit cost and transit trips for all block groups.

Constructing the H+T Index

Because the H+T Index was constructed to estimate the three dependent variables (auto ownership, auto use, and transit use) as functions of independent variables, any set of independent variables can be altered to see how the outputs are affected. To focus on the effects of the built environment, the independent household variables (income, household size, and commuters per household) were set at fixed values. This controls for any variation in the dependent variables that is a function of household characteristics, leaving the remaining variation a sole function of the built environment. In other words, by establishing and running the model for a "typical household," (one defined as earning the regional area median income, having the regional average household size, and having the regional average number of commuters per household) any variation observed in transportation costs is due to place and location, not household characteristics.

The Regional Typical Household considers all types of households in the region, and does not represent a specific household, but an average of all households. Every region has a unique mix of households: two-commuter households, single-earner households, adults with no children, single people, etc. - so the Regional Typical Household represents a composite of the broad range of households within a region.

Model Findings

The following three tables show the results of the regressions. This is the result of the process laid out above, where every independent variable is linearized using the function show in Table 3 above, and every combination that is significant and improves the R2 by at least 0.001%

Table 4: Results of Auto Ownership Regression (R^2 = 78.76% for Untransformed Variable)

Variable 1	Variable 2	Value	Est. Error	Individual R^2 of Transformed Variable	Incremental R^2 Transformed Variable
Intercept		1.10E-01	1.00E-02	NA	NA
Percent Rental	Renter Gravity	-2.12E-04	9.00E-06	44.50%	44.50%
Commuters	Percent Rental	3.40E-04	6.00E-05	14.90%	61%
TAS Jobs	Peak Service	-3.18E-09	9.00E-11	28.20%	67.70%
HH Size	Income	1.01E+04	3.31E+02	26.70%	71.90%
Job Mix Index (f)		8.20E-05	2.00E-06	36.20%	74.40%
SFD Fraction	Renter Gravity	2.39E-02	4.00E-04	26.90%	76.60%
Commuters	Block Size	-2.39E-02	7.00E-04	32.10%	77.90%
HH Gravity	Percent Rental	-6.90E-14	2.00E-15	31.30%	78.90%
SFD Fraction	Income	7.91E+02	1.36E+02	0%	79.30%
Job Gravity	SDF Gravity	1.74E-15	4.00E-17	13.90%	79.70%
SDF Gravity	Peak Service	-3.80E-12	1.00E-13	10.30%	80%
HH Size	Other TCI	-4.35E-05	8.00E-07	23.70%	80.20%
Income	Income	7.88E+07	1.24E+06	16%	80.40%
Income		-2.14E+04	3.03E+02	22.60%	80.60%
HH Size	Job Mix Index	-7.50E-05	2.00E-06	0.10%	80.90%
Block Size	Job Mix Index	1.60E-05	2.00E-07	23.10%	81%
Commuters	Income	5.06E+03	1.29E+02	7.10%	81.40%
Commuters	HH Density	-2.59E-03	7.00E-05	27.40%	81.50%
Commuters	Renter Gravity	1.12E-02	3.00E-04	0.20%	81.70%
Bus TCI	Job Mix Index	-6.00E-10	1.00E-10	26.70%	81.80%
Block Size	Block Size	-6.80E-03	2.00E-04	15.70%	81.80%
Block Size	Percent Rental	-3.80E-04	1.00E-05	9.30%	81.90%
Income	TAS Area	-2.00E+01	5.00E-01	31.30%	82%
Percent Rental	TAS Area	1.03E-05	3.00E-07	35.50%	82.10%
HH Gravity	SDF Gravity	-1.35E-20	4.00E-22	20.80%	82.20%
Bus TCI	HH Gravity	1.85E-16	7.00E-18	20.90%	82.20%
HH Size	Renter Gravity	-2.90E-02	1.00E-03	38.50%	82.20%
SDF Gravity	TAS Jobs	3.40E-16	1.00E-17	14.80%	82.30%
Job Mix Index	TAS Jobs	-3.80E-11	1.00E-12	24.30%	82.30%
Bus TCI	Percent Rental	-1.70E-07	8.00E-09	35.40%	82.40%
Peak Service	Peak Service	3.00E-05	2.00E-06	30.10%	82.40%
Block Size	TAS Jobs	4.40E-08	3.00E-09	29.10%	82.40%
HH Density	HH Density	2.70E-06	2.00E-07	2.70%	82.40%
Commuters	HH Size	1.09E-01	9.00E-03	0.70%	82.40%

Table 5: Results of Auto Use (VMT) Regression (R^2 = 83.23% for Untransformed Variable)

Variable 1	Variable 2	Value	Est. Error	Individual R^2 of Transformed Variable	Incremental R^2 Transformed Variable
Intercept		1.01E+01	1.00E-01	NA	NA
HH Size	HH Gravity	-2.90E-03	9.00E-04	68.20%	68.20%
Income	Percent Rental	-7.10E+01	1.00E+01	27.60%	79%
Commuters	Commuters	1.20E+00	3.00E-01	20%	80.40%
HH Density	Percent Rental	-3.30E-04	7.00E-05	57%	81.80%
Bus TCI	Peak Service	-4.00E-04	1.00E-04	61.40%	82.80%
Job Mix Index	SDF Gravity	1.20E-05	2.00E-06	46.40%	83.10%
Commuters	SDF Gravity	-9.00E-02	3.00E-02	27.40%	83.30%
Block Density	Bus TCI	-3.00E-03	1.00E-03	60%	83.40%
Commuters	Bus TCI	7.00E-03	2.00E-03	67.60%	83.50%
Block Density	Block Density	1.00E+00	1.00E-01	33.80%	83.60%
Commuters	HH Size	-4.50E+00	6.00E-01	30.60%	83.70%
Block Density	Renter Gravity	-8.00E-02	2.00E-02	0.30%	83.90%
sfd fraction	HH Gravity	5.00E-04	1.00E-04	7.10%	84%
sfd fraction	Other TCI	-2.80E-03	7.00E-04	10.70%	84%
Other TCI	Other TCI	3.50E-05	7.00E-06	42.20%	84.10%
HH Density	Other TCI	-4.40E-04	1.00E-04	47.20%	84.10%
Other TCI	Percent Rental	-3.20E-05	7.00E-06	55.80%	84.20%
HH Density	HH Density	7.00E-04	3.00E-04	20%	84.20%
Block Density	Other TCI	2.80E-03	1.00E-03	48%	84.20%
Block Density	Job Gravity	-6.00E-02	1.00E-02	8.60%	84.20%
Income	SDF Gravity	-1.51E+03	2.18E+02	17.40%	84.30%
HH Size	Income	3.37E+04	5.80E+03	23%	84.40%
Income	TAS Jobs	1.09E+03	1.88E+02	28%	84.40%
Income	Peak Service	-1.41E+03	2.80E+02	32.10%	84.40%
Job Mix Index	TAS Jobs	-6.00E-06	1.00E-06	4.70%	84.40%
Job Gravity	TAS Area	3.70E-03	8.00E-04	25.40%	84.60%
Commuters	TAS Jobs	-3.00E-02	1.00E-02	40.40%	84.60%
HH Size	TAS Area	6.00E-02	2.00E-02	33.70%	84.60%
Job Mix Index	HH Gravity	2.20E-07	6.00E-08	24.80%	84.60%
HH Gravity	HH Gravity	-2.00E-06	7.00E-07	53.50%	84.60%

Table 6: Results of Transit Use (R2 = 77.25% - for Untransformed Variable)

Variable		Value	Est. Error	Individual R^2 of Transformed Variable	Incremental R^2 Transformed Variable
Intercept		8.00E-01	1.00E-01	NA	NA
Job Mix Index	HH Gravity	1.79E-03	3.00E-05	64.90%	64.90%
Other TCI	Peak Service	-2.28E-02	9.00E-04	58.90%	69.80%
Peak Service	Peak Service	3.65E-02	8.00E-04	55.10%	70.90%
HH Gravity	TAS Area	-1.38E-07	4.00E-09	42.60%	71.50%
Job Mix Index	Job Mix Index	-8.90E-02	1.00E-03	51.80%	72.40%
Job Gravity	Peak Service	-8.10E-04	2.00E-05	49.60%	73.20%
Job Mix Index	Other TCI	2.81E-02	6.00E-04	40.80%	73.80%
Other TCI	TAS Jobs	4.10E-08	5.00E-09	37.10%	74.40%
Job Mix Index	Renter Gravity	-3.48E-06	7.00E-08	58.30%	74.70%
Job Mix Index	Income	3.17E+03	1.63E+02	0%	75.20%
Commuters	sfd fraction	5.90E-01	3.00E-02	22.70%	75.60%
HH Size	TAS Jobs	-1.65E-07	9.00E-09	42.20%	75.90%
Job Gravity	Renter Gravity	2.43E-08	7.00E-10	40.10%	76%
Block Size	SDF Gravity	-1.90E-10	1.00E-11	2.90%	76.10%
Commuters	HH Gravity	5.00E-04	2.00E-04	60.30%	76.20%
Job Mix Index	TAS Jobs	1.32E-07	5.00E-09	47.20%	76.30%
Renter Gravity	TAS Area	4.10E-10	1.00E-11	46.10%	76.40%
Renter Gravity	TAS Jobs	-5.00E-12	2.00E-13	28.40%	76.50%
Percent Rental	Peak Service	1.74E-03	8.00E-05	36.40%	76.60%
Commuters	Percent Rental	3.90E-03	9.00E-04	11.90%	76.60%
Income	Income	-1.43E+08	8.38E+06	0.30%	76.70%
Income	Other TCI	-1.64E+03	7.50E+01	20.50%	76.70%
Percent Rental	SDF Gravity	-6.80E-12	5.00E-13	5.40%	76.80%
SDF Gravity	Peak Service	1.70E-10	7.00E-12	8.40%	76.80%
Job Gravity	SDF Gravity	-3.70E-12	2.00E-13	11%	77.10%
gross hh density	SDF Gravity	9.40E-11	8.00E-12	16.10%	77.10%
Commuters	Income	-2.19E+04	1.24E+03	0.50%	77.20%
Percent Rental		-1.70E-02	1.00E-03	13.60%	77.20%
Job Gravity	Income	4.00E+00	2.00E+00	18.60%	77.20%
Bus TCI	Peak Service	-8.50E-04	8.00E-05	51.80%	77.20%
Commuters	Bus TCI	6.90E-03	7.00E-04	54.50%	77.20%
Income	TAS Area	3.30E-01	2.00E-02	7.10%	77.20%
Income	Percent Rental	2.22E+02	1.20E+01	3.70%	77.20%

Neighborhood Characteristic Scores

The H+T is based on the idea that some places are more efficient than others, a concept known as location efficiency. One way to measure this efficiency is to examine the extent to which a place is auto dependent. By looking at the place driven components of the regression equation to predict auto ownership (and in one case the transit use equation), comparisons between places can be made. Location efficiency can be scored by controlling for household characteristics and examining at how block groups compare with one another with regard to compact development, access to employment and variety of jobs, and level of transit service. Three scores were developed to make such comparisons: the Compact Neighborhood Score, Job Access Score and AllTransit™ Access Score. All are available on the H+T mapping tool, data download, and the H+T Fact Sheet.

They are all scores in the sense that they do not have a direct value of location efficiency to them, but are the rank of the block group relative to all other block groups in the H+T Index. This is accomplished by first evaluating the components of the equation of the subset of independent variables (for example, the Job Access Score uses Employment Access, and Job Mix Index), then this number (V_r) is scaled from 0 to 100 (I_r), and then all the block groups are ranked and given a number from 0 to 10 (S_{10}) reflecting their rank. The final score is one tenth of the percentile they fall into; a score of 5.5 for a particular block groups represents that that block group is in the 55th percentile of all block groups. The following equations show this calculation:

$$V_r = \sum_{i=1}^{n} C_i \times f_i(X_i)$$

Equation 9: Calculation of Generic Raw Value V_r

Where:

- i is the index or the variables used in this score
- n is the total number of variables used for this score
- C_i is the fit coefficient from the regression equation for the i^{th} variable
- X_i is the value of the i^{th} variable for this block group
- $f_i()$ is the linear transformation for the i^{th} variable

This value is then transformed into a number from 0 – 100 by using the same equations used in the Bus Access Index, the Rail Access Index and the Employment Mix Index, shown below:

$$I_r \equiv 100 \times \frac{V_r - V_{min}}{V_{max} - V_{min}}$$

Equation 10: Calculation of Generic Raw Index I_r

Where:

V_{min} is the minimum value for all block groups and
V_{max} is the maximum value for all block groups.

The value of this index is used then to rank all block groups (using a "dense ranking" where two block groups with the exact same value get the same rank, and the next one in gets the next rank) then this rank is turned into a number from 1 to 10 much as above:

$$S_{10} \equiv 10 \times \frac{R_r - R_{min}}{R_{max} - R_{min}}$$

Equation 11: Calculation of Generic Score S_{10}

Where:

R_r is the dense rank of the block group
R_{min} is the minimum dense rank (usually equal to one)
R_{max} is the maximum dense rank

This then gives the score which goes from 0 to 10.

The three scores use different inputs and regression equations, listed in Table 7 below.

Table 7: Neighborhood Characteristic Scores Definitions

Score	List of Independent Variables	Regression Equation
Compact Neighborhood Score	• Gross Household Density • Regional Household Intensity • Fraction of Single Family Detached Housing • Fraction of Rental Housing • Block size	Autos per Household
Job Access Score	• Employment Gravity • Employment Mix Index	Autos per Household
AllTransit™ Performance Score	• Transit Connectivity Index • TAS Jobs • Average Available Peak Transit Trips per Week	Percent Transit Journey to Work