H + T Index Methods

March 2015

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 to estimate three dependent variables (auto ownership, auto use, and transit use) as functions of 13 independent variables (median household income, average household size, average commuters per household, gross household density, Regional Household Intensity, fraction of single family detached housing, Employment Access Index, Employment Mix Index, block density, Transit Connectivity Index, Average Available Transit Trips per Week, Transit Access Shed, and Jobs within the Transit Access Shed). To hone in 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 "typical household" any variation observed in transportation costs is due to place and location, not household characteristics.

Significant Differences in the new Transportation Cost Model

Several improvements have been made to the H+T Index including a simplification of the transportation cost model and a new method to derive auto ownership costs. Previous versions of the H+T Index used a non-linear regression technique that made the model more difficult to understand. This version of the H+T Index uses the ordinary least square and simple variable transformations to accomplish the regression. The list of independent variables has also been simplified and excludes variables that are highly collinear so as to give greater significance to the remaining variables.

Another change to this version of the H+T Index is the use of a new method and data source for calculating the cost of owning and operating a vehicle (see Transportation Cost Calculation on page 25). Previous versions of the H+T Index used auto ownership costs from AAA's Your Driving Costs, one of the most widely used sources of costs for owning and operating a vehicle. It is employed by several government agencies, industry experts and transportation consultants and is based on amortized purchase costs for the first five years of a newly purchased auto. This version of the H+T Index uses costs derived from research commissioned and published by the Department of Transportation (DOT) and the Department of Housing and Urban Development (HUD) using the Consumer Expenditure Survey data . These costs were adjusted to reflect the 26% increase in CES auto ownership costs from year to 2013. This cost data reflects autos 10 years old or less, as compared with AAA estimated for cars five years old or less.

Geographic Level and Data Availability

The H+T Index was constructed at the Census block group level. Currently the H+T Index covers all 917 Metropolitan and Micropolitan Areas in the United States also known as Core Based Statistical Areas (CBSAs), as defined by the Office of Management and Budget (OMB) in 2013. Due to incompatible and insufficient data, all twelve regions in Puerto Rico were excluded.

Data Sources

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

- 2009-2013 American Community Survey 5-year Estimate (2013 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 2011 Census data are used here.
- 2000 Census Transportation Planning Package (2000 CTPP) –data on employment for census block groups are used in place of LEHD data for the state of Massachusetts (where LEHD data is not available).
- State of Massachusetts ES-202 - employment by industry category, used to leverage the 2000 CTPP employment data to 2013.
- Average annual expenditures and characteristics of all consumer units, from the Consumer Expenditure Survey, 2006-2012 and 2013, used to inflate the cost of auto ownership from the 2010 data above.
- 2013 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 2013 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 2013 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 in order to obtain the best fit possible from all of 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 (i.e. dollars per mile)
F = function of the independent variables (F_{AO} is auto ownership, F_{AU} is auto use, and F_{TU} is transit use)

Table 1: Independent Variables Overview

Variable	Description	Data Source	Type
Median HH Income	Median household Income in the block group	2013 ACS	Household
Commuters/HH	Workers per household who do not work at home	2013 ACS	Household
Avg. HH Size	Average number of people per household	2013 ACS	Household
Gross Household Density	Number of households divided by the land area in the Census block group	2013 ACS, TIGER/Line files	Neighborhood Characteristic (housing density)

Regional Household Intensity	Households summed divided by the distance squared in miles between block group by (the households in the block group are not included)	2013 ACS, TIGER/Line files	Neighborhood Characteristic (housing density)
Fraction of Single Family Detached Housing	Fraction of single family detached households in the block group	2013 ACS, TIGER/Line files	Neighborhood Characteristic (housing density)
Employment Access Index	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	Number of block per acre Weighted sum of 13 different employment types each scaled by a coefficient that are optimized using transit use	Census LEHD-LODES	Neighborhood Characteristic (employment)
Block Density	Number of block per acre	TIGER/Line files	Neighborhood Characteristic (walkability)
Transit Connectivity Index	Sum of buses/trains per week scaled by overlap of 1/8 mile rings about every stop that intersects the block group	CNT AllTransit	Neighborhood Characteristic (transit)
Average Available Transit Trips per Week	Number of possible transit rides within the block group and a ¼ mile of its border.	CNT AllTransit	Neighborhood Characteristic (transit)
Transit Access Shed	Total area that transit riders from 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 the TAS area	CNT AllTransit	Neighborhood Characteristic (transit)

Independent Variables: Household Characteristics

The 2013 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 2013 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 2013 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 2013 ACS. It is simply the number of households in a census block group divided by the area of land within the block group

Regional Household Intensity

The Regional 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 household 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 both the intensity of housing development in the region around the block group.

The Regional Household Intensity is calculated as:

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

Equation 2: Regional Household Intensity Definition

Where:

- H is the Regional 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 Regional 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, in order to expedite the calculation, the calculation uses the[1]:

- State totals when the state is not the same as the given block group and is more than 88 miles away,
- County totals when the county is not the same as the given block group and is more than 11.5 miles away, and
- Census tract totals when the tract is not the same at the given block group and is more than 2.5 miles away.

Fraction of Single Family Detached Households

The fraction of single family detached households is calculated using the 2013 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 density, 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 density 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 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.

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.

Massachusetts Employment Data

Employment data for the state of Massachusetts is not included in the LODES database. As noted on the LEHD website: “All 50 states, the District of Columbia, Puerto Rico, and the U.S. Virgin Islands have

[1] These distance thresholds were developed using the average distance between the geographic entities.

joined the LED Partnership, although the LEHD program is not yet producing public-use statistics for Massachusetts, Puerto Rico, or the U.S. Virgin Islands." [2]

A method was developed using Massachusetts ES202 data and the 2000 CTPP to estimate employment for this state. Employing the Massachusetts ES202 database query tool 2013 county employment was collected.[3] The 2000 CTPP was used to obtain the number of employees and related industry type. The constant share method was applied to the 2000 CTPP employment data at the block group level to estimate 2013 employment for every block group in Massachusetts. The ES202 data provided by the state also breaks up the employment into 13 categories by industry type; the industry sectors from the ES 202 were also extrapolated to 2013 using the 2000 CTPP and the constant share method.

Employment Access Index

The Employment Access Index 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 Access Index 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 Access Index is calculated as:

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

Equation 3: Employment Access Index 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 Access Index. 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:[4]

- State totals when the state is not the same as the given block group and is more than 88 miles away,
- County totals when the county is not the same as the given block group and is more than 11.5 miles away, and

[2] http://lehd.ces.census.gov/
[3] http://lmi2.detma.org/lmi/lmi_es_a.asp
[4] These distance thresholds were developed using the average distance between the geographic entities and factor determined such that the calculation remains consistent.

- Census tract totals when the tract is not the same at the given block group and is more than 2.5 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 fraction of one type of job (out of 13 total types) divided by all job 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 51.2%, but when the employment mix index is included the R^2 increases to 60.7%, a significant improvement.

Table 2 lists the 13 employment categories derived from the 2000 CTPP with the transformation function and the weight used. The variable is the fraction of the gravity measure of the jobs of the given type divided by overall employment access index (see above), i.e. the fraction of all accessible jobs of the given type. The weight is determined by regressing all of the other independent variables and these 13 against percent transit use for journey to work. The transformation function column displays the function used as the linear transformation (Linear (x), Square Root ($\sqrt{x}$), Natural Log ($\ln(x)$)), and Inverse ($1/x$)). The *Index Effect* column indicates what happens to the value of the Employment Mix Index when the fraction of the given employment type increases.

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

Category	Variable Name	Transformation Function	Weight	Index Effect
Agriculture, Fishing, Forestry, Mining	emp_f02	$1/x$	0.00194	Reduce
Construction	emp_f03	x	15.7	Increase
Manufacturing	emp_f04	$\ln(x)$	-.77	Reduce
Wholesale	emp_f05	x	-21	Reduce
Retail	emp_f06	$\ln(x)$	-2.99	Reduce
Transportation, Warehousing, Utilities	emp_f07	$1/x$	-.0358	Increase
Information	emp_f08	$1/x$	-.0034	Increase
FIRE	emp_f09	x	5.9	Increase
Professional ,STEM, Mgmt., Admin, Waste	emp_f10	x	-5.4	Reduce
Education, Health, Social	emp_f11	x	3.2	Increase
Arts, Entertainment, Rec, Accommodation, Food	emp_f12	x	-5.2	Reduce
Other Services	emp_f13	$1/x$	-.009	Increase
Public Administration	emp_f14	$1/x$	-.039	Increase

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
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.

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

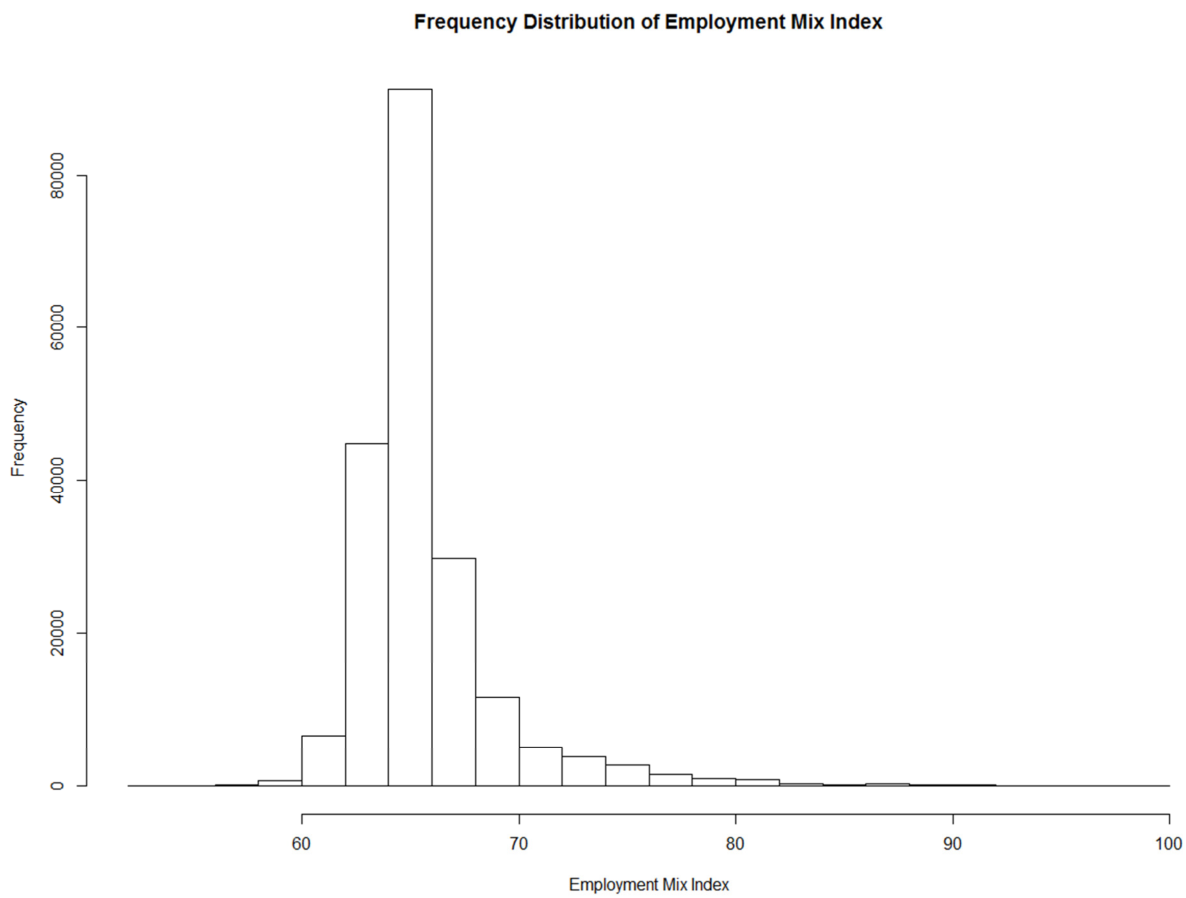
Equation 5: Definition of Employment Mix Index

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 is calculated of all Census block groups in the country; the following graph shows the distribution of values for this index for only the Census block groups in the sample used the H+T index.

Figure 1: Frequency Distribution of Employment Mix Index for Census Block Groups in CBSAs



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 proprietary dataset known as All Transit™. All Transit is an online tool that facilitates the collection, normalization, aggregation, and analysis of GTFS data to determine fixed-route transit service.

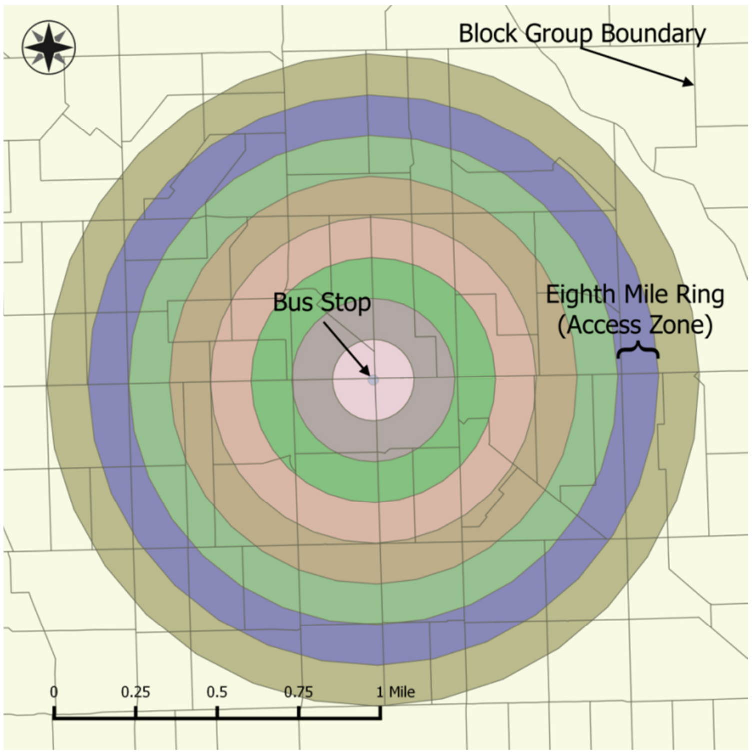
To date, CNT has compiled stop, station, and frequency data for bus, rail, and ferry service for all major transit agencies in regions with populations greater than 250,000 (with the exception of Dayton, OH; Roanoke, VA; and York-Hanover, PA). Attachment A lists the transit agencies for which data has been compiled. In regions where data is not available, CNT has constructed household transportation behavior models, that while less robust are still very good. Developing this set of models for regions where good transit data is not available allows for coverage of all the CBSAs.

Four measures of transit access are used in the model: the Transit Connectivity Index (TCI), Transit Access Shed (TAS), Transit Access Shed Jobs (TAS Jobs), and Average Available Transit Trips per Week. The TCI estimates of how many transit opportunities are within walking distance of a census block group. The TAS is a proxy measure for how far one can travel in 30 minutes on transit, while the TAS Jobs is the sum of the total number of jobs within the TAS.

Transit Connectivity Index

The Transit Connectivity Index is a measure of access to bus stops that CNT developed specifically for use in the household transportation cost model. To calculate this measure, eight concentric rings one-eighth of a mile in width (access zones) were plotted around each bus stop (see **Error! Reference source not found.**).

Figure 2: Illustration of 8 1/8th mile Rings around Transit Stop (For a Single Bus Stop in a Chicago Neighborhood)



Using these access zones, the following are defined for each block group:

Table 3: Bus Transit Connectivity Variables

Variable	Description
LC	Land area of the block group covered by access zone
SFV	Service frequency value
BLA	Total block group land area

At each block group, eight transit access values were calculated for each bus stop where at least one of the access zones intersects the block group. The following formula is used for each bus stop to obtain the scaled frequency (SF) by ring for each block group.

$$SF_d = \sum_{i=1}^{n} \frac{LC_{i,d} SFV_{i,d}}{BLA}$$

Equation 6: Bus SF_d Calculation

Where:

d is the index across the eight concentric circular access zones,

bgi is the index the bus stops that have access zones ring d that intersect the block group

n is the total intersecting bus stop access zone ring d.

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 bus stops.

The farther an access zone is from its transit node, the less of a contribution it should make to the level of access in any block group it intersects, however the relative area covered by these distant access zones is larger because of their shape. In order to account for the decreasing access benefits at greater distances and the increased area coverage a weight is given to each value of SF_d calculated using regression. Measured values of percent journey to work by transit were regressed against the 8 SF_d values (as defined above) using an ordinary least square to define the weight of each of the eight rings.

The sum of the weights times the SD_d are calculated for each block group. This quantity is not easily translated since it is the combination of many factors, so the final value for this index is a number from 0-100 representing the first stage of a fit for the use of transit for commuter's journey to work by the following formula:

$$Bus\ TCI \equiv 100 \times \frac{STD - STD_{min}}{STD_{max} - STD_{min}}$$

Equation 7: Bus Connectivity Index Calculation

.

Where:

STD is the sum of all of the SF_d i.e. $STD = \sum_{d=1}^{8} Wt_d SF_d$,

STD_{min} is the minimum value for all block groups and

STD_{max} is the maximum value for all block groups.

The same method was used to run the Transit Connectivity Index for other modes of transit (commuter rail, subway, metro, tram, streetcar, light rail, ferry service, gondola, funicular, and cable car); however these modes use sixteen instead of eight one-eighth mile rings (2 miles). These components of the Transit Connectivity Index are added to the similar bus component to make the final Transit Connectivity Index. The following table shows the final weighting for the ring (both bus and other) that create the final Transit Connectivity Index.

Table 4: Coefficients use in Transit Connectivity Index (Note that missing rings indicate no statistical significance and/or covariant)

Description	Variable Name	Transformation Function	Weight	Index Effect
Bus Ring 1	bus_ring01	$\sqrt{x}$	.107	Increase
Bus Ring 5	bus_ring05	x	0.00003	Increase
Bus Ring 6	bus_ring06	x	0.00003	Increase
Bus Ring 8	bus_ring08	x	0.000061	Increase
Other Ring 1	other_ring01	x	0.00054	Increase
Other Ring 2	other_ring02	$\ln(1+x)$	0.50	Increase
Other Ring 4	other_ring04	$\ln(1+x)$	0.28	Increase
Other Ring 5	other_ring05	$1/(1+x)$	-1.0	Increase
Other Ring 9	other_ring09	$\ln(1+x)$	0.12	Increase
Other Ring 11	other_ring11	$1/(1+x)$	-0.92	Increase
Other Ring 15	other_ring15	x	0.00004	Increase
Other Ring 16	other_ring16	$\ln(1+x)$	0.21	Increase

Transit Access Shed

The Transit Access Shed (TAS) is defined as a geographic area accessible within 30 minutes by public transportation. This measure was derived from the All Transit GTFS data. For each transit stop, all stops that can be reached within 30 minutes were identified. One transfer within 600 meters 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. In order 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 and the 2000 CTPP for the state of Massachusetts.

Average Available 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.

Dependent Variables

Auto Ownership

For the dependent variable auto ownership, the regression analysis was fit using measured data on auto ownership obtained from the 2013 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.

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 2010 through 2012 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.

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.[5]

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

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 2013 ACS, Means of transportation to work was used to calculate a percent of commuters utilizing public transit.

Household Transportation Regression Analysis

For this version of the H+T Index the model has been simplified. The most important relationships between independent and dependent variables are non-linear, so in previous iterations of the household transportation model a non-linear regression was used. In this iteration this non-linearity is compensated for by using simple transformation functions. These functions (Linear (x), Square Root $(\sqrt{x})$, Natural Log $(\ln(x))$, and Inverse $(1/x)$) 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. A model is constructed for each dependent variable (auto ownership, auto use and transit use) using the four transit variables (Bus Access Index, Rail Access Index, Transit Shed and Jobs in Transit Shed) where these measures are available. Another model is also constructed leaving these transit measures out in order to make a good model for regions where transit data is missing.

An ordinary least square regression was used to estimate the fit coefficients; the equation that will estimate the dependent variable from independent variables is:

$$D = I + \sum_{1=1}^{n} C_i \times f_i(x_i)$$

Equation 8: Equation for Estimating Dependent Variable from Regression Coefficients and Independent Variables

Where:

D is the 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
C_i is the fit coefficient for the i^{th} independent variable
f_i is the linearization transformation function
x_i is the value of the i^{th} independent variable

Choosing Linear Transformation Functions for the Independent Variables

In order to address the nonlinear nature of the relationships between the independent and the dependent variables, a linear transformation function was chosen. These were limited to Linear (x), Square Root ($\sqrt{x}$), Natural Log ($\ln(x)$), and Inverse ($^1/_x$). For variables that could have a legitimate value of zero, the functions Safe Natural Log $\ln(x+1)$ and Safe Inverse $^1/_{(x+1)}$ were also used. In order to choose the best transformation each variable was tested to determine which transformation resulted in the best fit.

The example below considers how gross household density drives auto ownership. Figure 4 and Figure 5 show that the relationship between auto ownership and gross household density is nonlinear. However, the relationship between auto ownership and the natural log of the gross household density shows a more linear relationship; note the increase in R^2 by almost eight percentage points. This technique was then repeated for every variable for every model to select the optimal transformation.

Figure 3: Fit vs. Median Household Income

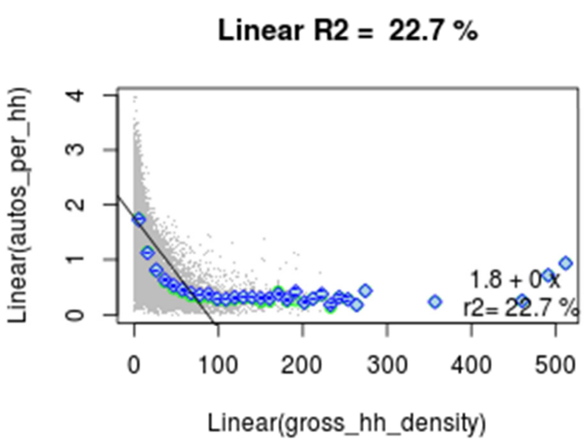
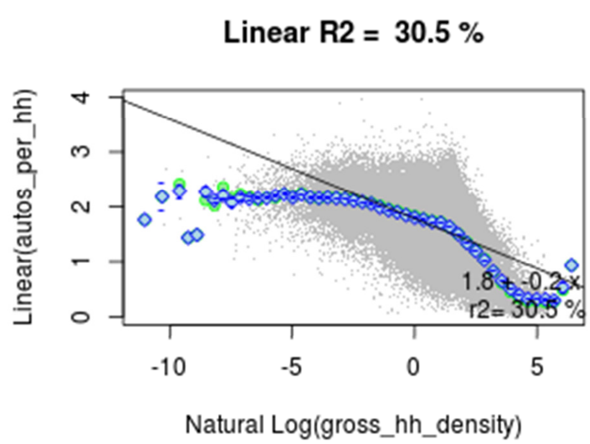


Figure 4: Fit vs. the Natural Log of the Median Household Income



Choosing Independent Variables

In order to test the statistical significance of variables and to determine those that would reduce the multicolinearity of the set of independent variables, an initial set of variables from previous versions of H+T Index models were examined. Table 5 below lists the result of a regression analysis using all Census Block Groups in CBSAs that have non suppressed data. They are listed in the order that they contribute to the overall R^2. The *VIF* column gives the variance inflation factor (VIF) for multicollinearity; a value below 10 is generally thought to be an acceptable level for multicollinearity. Note that in Figure 6 below shows the diminishing return on including all of the variables, and in this and all of the six models used, variables that added less than 0.1% to the R^2 were not included in the final fit. Table 5 below shows the final variables used in the autos per household model. All of the VIFs are 5 or less; this is true for all six models.

Table 5: Summary of all variables examined Autos per Household Regression [6]

[6] In

Table 5: Summary of all variables examined Autos per Household Regression

Rank	Variable	Individual R2	Incremental R2	Change in R2	VIF
1	Fraction of Single Family Detached Housing	49.2%	49.2%	NA	1.9
2	Commuters per Household	22.9%	59.5%	10.3%	1.8
3	Transit Connectivity Index	35.1%	69%	9.5%	5.7
4	Median Household Income	29.9%	75.1%	6.2%	1.6
5	Gross Household Density	30.5%	77.7%	2.6%	5.2
6	Employment Mix	29%	79.1%	1.4%	2.4
7	Household Size	12.9%	80%	0.9%	1.6
8	Regional Household Intensity	28.6%	80.3%	0.4%	7.6
9	Block Density	24.4%	80.4%	0.1%	3.7
10	Employment Gravity	5.8%	80.5%	0.1%	1.5
11	TAS Jobs	23.9%	80.5%	0%	4.9
12	TAS	16.5%	80.6%	0%	2
13	Average Available Transit Trips per Week	30.5%	80.6%	0%	3.9

and Table 6 the color of the row highlights the degree to which the variables increases the R^2,each color represents a range:

- Gold variables change the R^2 by more than 5%

Rank	Variable	Individual R^2	Incremental R^2	Change in R^2	VIF
1	Fraction of Single Family Detached Housing	49.2%	49.2%	NA	1.9
2	Commuters per Household	22.9%	59.5%	10.3%	1.8
3	Transit Connectivity Index	35.1%	69%	9.5%	5.7
4	Median Household Income	29.9%	75.1%	6.2%	1.6
5	Gross Household Density	30.5%	77.7%	2.6%	5.2
6	Employment Mix	29%	79.1%	1.4%	2.4
7	Household Size	12.9%	80%	0.9%	1.6
8	Regional Household Intensity	28.6%	80.3%	0.4%	7.6
9	Block Density	24.4%	80.4%	0.1%	3.7
10	Employment Gravity	5.8%	80.5%	0.1%	1.5
11	TAS Jobs	23.9%	80.5%	0%	4.9
12	TAS	16.5%	80.6%	0%	2
13	Average Available Transit Trips per Week	30.5%	80.6%	0%	3.9

- Green variables change the R^2 from 1% to 5%
- Kaki variables change the R^2 from 0.5% to 1%
- Salmon variables change the R^2 from 0.1% to 0.5%
- Tomato Red variables change the R^2 less than 0.1%

Figure 5: Graph of R2 as More Independent Variables are added to the Regression for Autos per Household

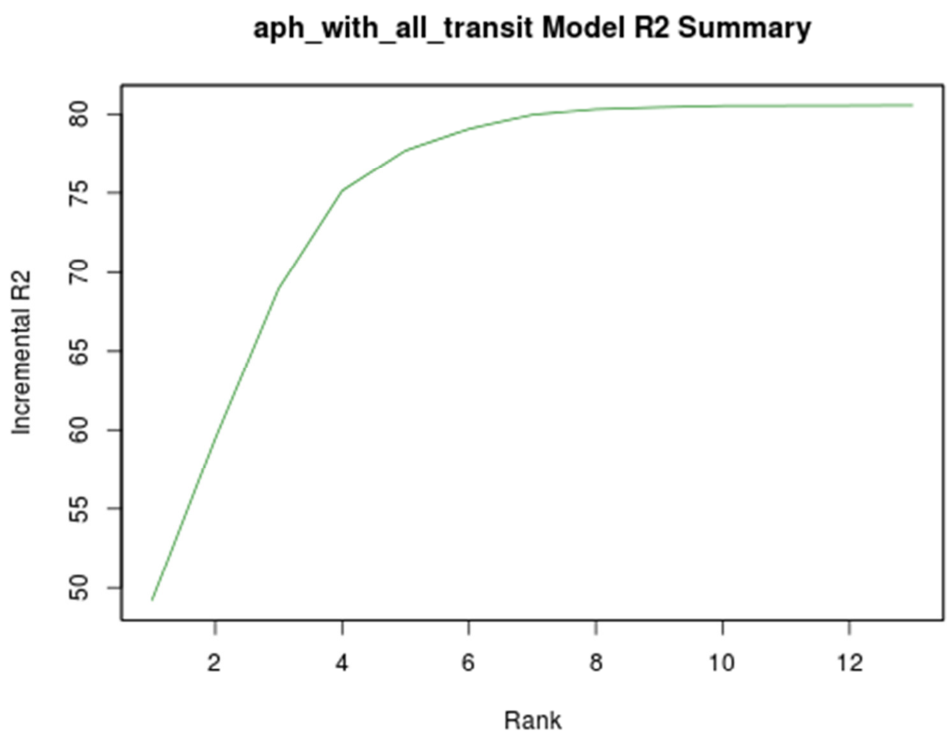


Table 6: Final Variables Included in the Autos per Household Regression

Rank	Variable	Individual R^2	Incremental R^2	Change in R^2	VIF
1	Fraction of Single Family Detached Housing	49.2%	49.2%	NA	1.9
2	Commuters per Household	22.9%	59.5%	10.3%	1.8
3	Transit Connectivity Index	35.1%	69%	9.5%	4.2
4	Median Household Income	29.9%	75.1%	6.2%	1.6
5	Gross Household Density	30.5%	77.7%	2.6%	5.2
6	Employment Mix	29%	79.1%	1.4%	2.4
7	Household Size	12.9%	80%	0.9%	1.6
8	Regional Household Intensity	28.6%	80.3%	0.4%	4.3
9	Block Density	24.4%	80.4%	0.1%	3.7
10	Employment Gravity	5.8%	80.5%	0.1%	1.5

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 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 2013 dollars. The factor used is derived from the CES; the average expenditure in 2010 is $2,588 and in 2013 it is $3,271, thus the factor applied is 1.26.

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.

Transit Use Costs

The 2013 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 2013 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 located 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 2013 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 a number of counties for which GTFS data are not available and/or there was no revenue listed in the 2013 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 end 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. In order 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 takes into account 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 six tables show the results of the six regressions. The *Function* column indicates what linearization function was used, the *Value* column give the value of the fit coefficient, the *Error* column gives the value of the standard error on the coefficient and the *Trend* column show what direction the dependent variable will change when the independent variable is increased.

Table 7: Results of Auto Ownership Regression including the Transit Measures in the Fit

Model	Auto Ownership with Transit			
R^2 80.6%				
Variable	**Function**	**Value**	**Error**	**Trend**
Fraction of Single Family Detached Housing	x	0.390	0.003	Increase
Commuters/Household	1/(1+x)	-1.27	0.01	Increase
Transit Connectivity Index	x	-0.0055	0.0001	Decrease
Median Household Income	$\sqrt{x}$	0.00251	0.00001	Increase
Gross Household Density	ln(x)	-0.0327	0.0009	Decrease
Employment Mix Index	x	-0.0183	0.0003	Decrease
Avg. Household Size	x	0.098	0.001	Increase
Regional Household Intensity	x	-0.00000161	0.00000003	Decrease
Block Density	ln(x)	-0.028	0.001	Decrease
Employment Access Index	1/x	110	4	Decrease
Intercept	NA	2.42	0.02	NA

Table 8: Results of Auto Ownership Regression Excluding the Transit Measures in the Fit

Model	Auto Ownership without Transit			
R^2 79.0%				
Variable	**Function**	**Value**	**Error**	**Trend**
Fraction of Single Family Detached Housing	x	0.387	0.002	Increase
Median Household Income	ln(x)	0.303	0.001	Increase
Regional Household Intensity	x	-0.00000236	0.00000002	Decrease
Commuters/Household	ln(1+x)	0.600	0.005	Increase
Gross Household Density	ln(x)	-0.0584	0.0005	Decrease
Employment Mix Index	x	-0.0190	0.0002	Decrease
Avg. Household Size	x	0.095	0.001	Increase
Block Density	$\sqrt{x}$	-0.155	0.005	Decrease
Intercept	NA	-1.14	0.02	NA

Table 9: Results of Auto Use (VMT) Regression Including the Transit Measures in the Fit

Model	Auto Use with Transit			
R^2 82.3%				
Variable	**Function**	**Value**	**Error**	**Trend**
Fraction of Single Family Detached Housing	x	3905	158	Increase
Average Available Transit Trips per Week	x	-1.8	0.2	Decrease
Commuters/Household	x	4152	162	Increase
Gross Household Density	1/(1+x)	3822	271	Decrease
Regional Household Intensity	x	-0.046	0.002	Decrease
Transit Connectivity Index	1/(1+x)	1870	221	Decrease

Median Household Income	1/x	-64375203	3612868	Increase
Avg. Household Size	$\sqrt{\ln(x)x}$	3182	199	Increase
Employment Access Index	1/x	7398057	1078476	Decrease
Transit Access Shed	$\sqrt{x}$	-0.09	0.01	Decrease
Intercept	NA	9577	279	NA

Table 10: Results of Auto Use (VMT) Regression Excluding the Transit Measures in the Fit

Model	**Auto Use without Transit**			
R^2 790.8%				
Variable	**Function**	**Value**	**Error**	**Trend**
Regional Household Intensity	$\sqrt{x}$	-27	1	Decrease
Commuters/Household	x	4386	160	Increase
Fraction of Single Family Detached Housing	ln(1+x)	5668	256	Increase
Block Density	$\ln(x)$	-911	65	Decrease
Median Household Income	$\ln(x)$	2020	103	Increase
Avg. Household Size	$\ln(x)$	3602	210	Increase
Employment Access Index	(x)	-621	92	Decrease
Intercept	NA	-6218	1168	NA

Table 11: Results of Transit Use Regression Including the Transit Measures in the Fit

Model	**Transit with Transit**			
R^2 74.7%				
Variable	**Function**	**Value**	**Error**	**Trend**
Regional Household Intensity	x	0.000158	0.000001	Increase
Transit Connectivity	x	0.467	0.004	Increase
Employment Access Index	x	-0.0000535	0.0000005	Decrease
Employment Mix Index	x	0.756	0.008	Increase
Fraction of Single Family Detached Housing	$\sqrt{x}$	-3.66	0.10	Decrease
Transit Access Shed	x	-0.0000000264	0.0000000004	Decrease
Transit Access Shed Jobs	x	0.0000058	0.0000002	Increase
Median Household Income	1/x	37674	1385	Decrease
Average Available Transit Trips per Week	x	0.0041	0.0002	Increase
Avg. Household Size	1/x	-5.2	0.2	Increase
Intercept	NA	-44.2	0.5	NA

Table 12: Results of Transit Use Regression Excluding the Transit Measures in the Fit

Model	**Transit without Transit**			
R^2 72.1%				
Variable	**Function**	**Value**	**Error**	**Trend**
Regional Household Intensity	x	0.0002735	0.0000008	Increase

Employment Mix Index	$1/x$	-4476	32	Increase
Employment Access Index	x	-0.0000447	0.0000004	Decrease
Fraction of Single Family Detached Housing	$\ln(1+x)$	-5.79	0.08	Decrease
Gross Household Density	1/(1+x)	2.01	0.06	Decrease
Median Household Income	1/x	30178	1154	Decrease
Intercept	NA	71.9	0.5	NA

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 as 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 Transit 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 ith variable
X_i is the value of the ith variable for this block group
$f_i()$ is the linear transformation for the ith 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 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 13 below.

Table 13: 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 Households • Block Density	Autos per Household
Job Access Score	• Employment Gravity • Employment Mix Index	Autos per Household
Transit Access Score	• Transit Connectivity Index • TAS Sq. Meters • TAS Jobs • Average Available Transit Trips per Week	Percent Transit Journey to Work