

REVIEW OF THE CENTER FOR NEIGHBORHOOD TECHNOLOGY'S HOUSING AND TRANSPORTATION AFFORDABILITY INDEX

Report Submitted To:
Manhattan Strategy Group
708 Third Avenue, 5th Floor
New York, NY 10017

Report Submitted By:
Econsult Corporation
1435 Walnut St., Suite 300
Philadelphia PA 19102

Penn Institute for Urban Research
3733 Spruce St., Vance Hall 430
Philadelphia PA 19104

February 2012

TABLE OF CONTENTS

1.0	Introduction.....	1
2.0	Description of The Index.....	3
2.1	The Housing + Transportation Affordability Index	3
2.2	Housing	3
2.3	Transportation	4
2.4	H+T Index	9
3.0	Analysis of the Rigor of the H+T Index: Modelling Transportation Uses.....	10
3.1	Perspectives on the Transportation Component of the H+T Index.....	10
3.2	Regression Framework for Transportation Models.....	15
3.3	Predictions of the Transportation Component	24
3.4	Case Study: Philadelphia Data.....	25
4.0	Existing Housing and Transportation Indexes	28
4.1	Analysis of Approaches to Indexes	28
4.2	Alternative Measures of Housing and Transportation Affordability	30
4.3	H+T Index from the Perspective of Users.....	37
5.0	Results	39
5.1	Strengths and weaknesses	39
5.2	Recommendations	41
5.3	Conclusion	43
Appendix 1	Index theory.....	46
Appendix 2	Survey Responses	50

1.0 INTRODUCTION

The U.S. Department of Housing and Urban Development's (HUD) Office of Housing and Sustainable Communities (OSHC) is currently working on the development of a "Housing and Transportation Affordability Index" (HTA Index) and related tools to measure the combined cost of housing and transportation in metropolitan and micropolitan areas nationwide. The launching point for this proposed HTA Index is the H+T® Affordability Index developed by the Center for Neighborhood Technology (CNT).

The Manhattan Strategy Group asked the Penn Institute for Urban Research and Econsult Corporation (Econsult) to conduct a review of the H+T Index and of currently existing metrics to measure housing and transportation costs and assess their strength and weaknesses with "the goal of informing the development of HUD's HTA Index."¹

The overarching objective of this initiative is "strengthening linkage between housing and transportation investments and exploring how combined housing transportation costs could be used as a measure of affordability."²

Key objectives of the proposal relevant to the review include:

- "Increase the understanding of the impact of combined household housing and transportation costs on families and communities by establishing a baseline for these costs by location.
- Increase access to information about the combined cost of housing and transportation by location for consumers and the financial community through development of a Housing and Transportation Affordability Index (HTA Index) and a Housing and Transportation Cost Calculator.
- Analyze the impact of proposed regional and local land use investments in housing and employment investments on housing and transportation costs within metropolitan areas.
- Analyze the impacts of HUD programs on the combined cost of housing and transportation in HUD-supported communities, and assess the feasibility of improving outcomes by integrating the HTA Index as a measure of affordability into HUD's programs and policies."³

Our review assesses the rigor of the H+T Affordability Index. We examine the measures of housing costs and transportation costs used in the H+T Index but as mentioned in the RFP since "housing costs are less complicated and easier to model than transportation costs, the review devotes more effort to analyzing the

¹ Extracted from the RFP.

² Ibid.

³ Ibid.

methodology behind transportation cost modeling.⁴ In particular, we examine conceptual challenges inherent in measuring housing and transportation costs, the empirical validity of the data, and the statistical models and formula used to calculate the index. To assess the rigor of the H+T index, we replicate CNT's empirical findings, explore the alternative models, and examine patterns in their predictions. The impacts of temporal change are examined by comparing CNT's housing cost patterns with sales transactions data for Philadelphia over the last decade to see if there were systematic changes over time that are not captured by the H+T index.

We also conduct an assessment of the existing Housing and Transportation Indices available and of their conceptual grounding in the literature. To augment this part of the review we conduct a survey of current users of the H+T Index. The respondents' experiences with the Index helped us assess the Index's uses and suitability for its proposed applications, including informing housing and transportation policies and programs. Finally, we provide a preliminary assessment of the strengths and weaknesses of the approach adopted by CNT's H+T Index, and suggest modifications to the H+T Index to help it attain the goals of the proposed HTA Index.

⁴ Ibid.

2.0 DESCRIPTION OF THE INDEX

2.1 THE HOUSING + TRANSPORTATION AFFORDABILITY INDEX

The H&T index is composed of two main parts: housing costs and transportation costs. The geography of the index is Census block level detail covering all 940 Metropolitan Statistical Areas, excluding Puerto Rico. There are a variety of datasets and modeling procedures that go into the transportation portion of the index, while the housing portion is fairly straightforward. The difference between how housing and transportation are each constructed is due to the desire to control for the effect of demographics on transportation demand but not housing demand.

2.2 HOUSING

The aim of the housing cost index is to measure average housing expenditures at the block group level as a percent of metropolitan median income. The database used for housing costs is the 2009 American Community Survey 5-year estimates. There is no temporal variation in the housing cost index. The housing cost for a given block group is estimated as:

$$\text{Housing cost} = (OOHU * SMOC + ROHU * GR) / OHU$$

Where the variables are defined as:

Table 1: Housing Cost Variables

Variable	Description
OOHU	Owner Occupied Housing Units
SMOC	Selected Monthly Ownership Costs
ROHU	Renter Occupied Housing Units
GR	Gross Rent
OHU	Occupied Housing Units

The selected monthly ownership costs variable includes utilities, condo fees, property taxes, insurance, first and second mortgage payments, and payments for home equity loans. In short, the housing cost is simply the average out of pocket expenses for housing. This is also known as the “payments approach” or “cash approach” in the literature on housing cost indexes.

The final transformation of the data for the housing cost index is to estimate the average housing cost as a percentage of household income for a representative “control household”. There are three different representative “control households” and corresponding median household income variables used. These are shown in Table 2 below.

Table 2: Control Households

Control Household Type	Median Household Income Level
Regional Typical Household	CBSA
Regional Moderate Household	80% of CBSA
National Typical Household	National

The block group level household costs estimated in 1) above are divided by one of the three control household income levels to produce the final household cost portion of the index.

2.3 TRANSPORTATION

The transportation cost portion of the index is not as transparent or simple as the housing portion. A variety intermediary of datasets, models, and estimations are used to arrive at a final number. The conceptual goal of the transit index is to estimate what transportation costs would be for a representative “control household” in each census block group. In order to do this, the amount and type of transportation that would be consumed by a household with a particular set of demographics must be estimated. Then, as with the housing cost model, the cost will be divided by the “control household” income level to estimate total transportation cost as a percent of income.

2.3.1 THE GENERAL TRANSIT COST MODEL

At the highest level, the index can be broken into three parts: auto ownership costs, auto use costs, and transit use costs. The following function shows how these are estimated from components:

$$\text{Transportation Costs} = C_{AO}Q_{AO} + C_{AU}Q_{AU} + C_{TU}Q_{TU}$$

Where C_x is the cost per unit and Q_x is the utilization amount for the three transportation cost categories: AO is auto ownership, AU is auto use, and TU is transit use.

A step that is necessary in order to eventually control for demographics is to estimate Q_x as a function F_x of the household (H) and environmental (E) independent variables. Thus the total transportation cost becomes:

$$\text{Transit Costs} = C_{AO}F_{AO}(H,E) + C_{AU}F_{AU}(H,E) + C_{TU}F_{TU}(H,E)$$

Where H represents household specific independent variables, such as household size, E represents environment specific independent variables, such as level of transit service. Then, for instance, $C_{AU} * F_{AU}(H,E)$ represents the cost of auto use per mile times the number of auto miles driven, the latter of which is a function of household and environmental independent variables.

2.3.2 TRANSPORTATION UTILIZATION REGRESSION MODELS

For each of the three utilization variables a non-linear regression model of the following rational functional form is estimated:

$$\hat{F}_z = \sum_{j=1}^m \frac{\alpha_{1j}x_j + \alpha_{3j}x_j^2 + \alpha_{5j}x_j^3}{\alpha_{2j}x_j + \alpha_{4j}x_j^2 + \alpha_{6j}x_j^3}$$

Where x_1, x_2, \dots, x_m are the independent variables in Table 3 below, and z indexes Auto Ownership (AO), Auto Utilization (AU), and Transit Utilization (TU), the three utilization categories.

The environmental and household variables used in the regressions include both raw data from the ACS as well as calculated data. The following table summarizes each variable and its source.

Table 3: Independent Variables

Description	Type	Source
Average household size	Household	ACS average household size
Commuters per household	Household	household pop * (workers – workers at home) / (households*pop)
Median household income	Household	ACS median household income
Per Capita Household Income	Household	Median household income / average household size
Gross Household Density	Environment	Households/land area of block group
Transit Connectivity Index	Environment	Calculation, see below
Transit Access Shed	Environment	Calculation, see below
Intersection density	Environment	Number of intersections / block group area
Block group size	Environment	TIGER
Employment access index	Environment	Calculation, see below
Net Residential density	Environment	Calculation, see below

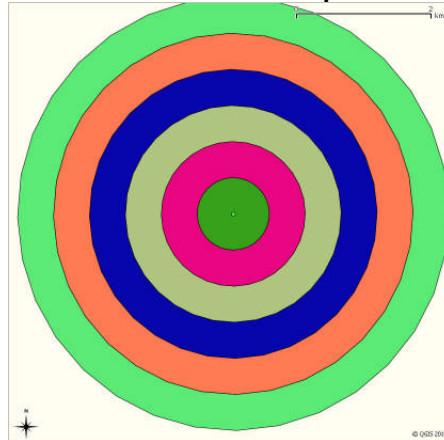
After the transportation regression models are run, the transit costs for a “control household” are estimated for each block group. This “control household” is defined with each household variable in Table 3 above being held constant using a regional average.

The calculated environmental variables are described in sections 2.3.3 through 2.3.6 below. Section 2.3.7 discusses the estimation of transit costs (C_{AO} , C_{AU} , and C_{TU}).

2.3.3 TRANSIT CONNECTIVITY INDEX

The transit connectivity index is a proprietary measure created by CNT. The index begins with a map of transit stops, and a buffer of concentric circles around each stop defining the “access area”. The circles are at quarter mile intervals for bus stops, and at half-mile intervals for other transit stops (Rail, Subway, Metro, Ferry, Cable car, Light rail, etc.). Figure 1 below, from CNT, illustrates:

**Figure 1: Access Area
Around Bus Stop**



Next, the following are defined for each block group:

Table 4: Transit Connectivity Variables

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

At each block group, six transit access values were calculated using the following formula:

$$TCI_{d,bg} = \frac{LC_{d,bg}SFV_{d,bg}}{BLA_{bg}}$$

Where d indexes across the six concentric circles, and bg indexes block group.

The six optimal weighting multipliers (W_d) are estimated by taking the average coefficients from regressing the six transit access variables (TCI_d) on two measures of transit utilization: autos per household (F_{AO}), and percent journey to work by transit (F_{TU}). Table 5 shows the weights for the six access areas.

Table 5: Transit Access Weights

Area	Bus Distance	Transit Distance	Weight
1	0.00 to 0.25 miles	0.0 to 0.5 miles	1.00
2	0.25 to 0.50 miles	0.5 to 1.0 miles	0.72
3	0.50 to 0.75 miles	1.0 to 1.5 miles	0.22
4	0.75 to 1.00 miles	1.5 to 2.0 miles	0.22
5	1.00 to 1.25 miles	2.0 to 2.5 miles	0.18
6	1.25 to 1.50 miles	2.5 to 3.0 miles	0.05

The final TCI estimate for each block group is taken as the sum of the six weighted transit access values ($TC_{d,bg}$) in each block group:

$$TCI_{bg} = \sum_{d=1}^6 \frac{W_d LC_{d,bg} SFV_{d,bg}}{BLA_{bg}}$$

One way to improve the transit connectivity measure would be to use network distances rather than straight line distance. This would reflect how far one has to walk, in contrast to straight line distances. However, this would likely improve measurement only slightly if at all. In its current form this is a detailed and useful measure of transit access.

2.3.4 TRANSIT ACCESS SHED

This measure of transit access starts by generating quarter mile buffers around all transit stops. Next, for each block group all of the transit stops within 30 minutes and one transfer of are located, and then the total land area within the quarter mile buffers of those stops is summed. This provides a measure for each block group of the total amount of area that is “easily accessible” using transit.

Here again one might improve the measure by instead looking at the land area within a quarter miles of walking from each transit stop. This would add extra weight to neighborhoods that are walkable. As with the transit connectivity index, it is likely that this measure is still very useful as it is, and that these improvements might not make a significant difference. Nevertheless, it is an issue worth exploring.

2.3.5 EMPLOYMENT ACCESS INDEX

The employment access index is intended to measure the availability of work near a block group. The underlying dataset is the 2007 Local Employment Dynamics data from the Census, which provides information on local labor market conditions at the block group level⁵. From this dataset it is known how many jobs there are in every block group. Then for each block group the following gravity model is used to weight jobs in nearby block groups:

⁵ Employment numbers are imputed for New Hampshire, Massachusetts, and D.C. using a constant share method to scale up 2000 tract level data from the Census Transportation Planning Package (CTPP) by 2007 county level BLS employment data.

$$E_{bg} = \sum_{i=1}^n \frac{p_i}{r_i^2}$$

Where p_i is the number of jobs in the nearby block group and r_i is the distance between the centroid of the given block group bg and the i th nearby block group. This is summed over the n block groups within 63 miles⁶.

One limitation of this data is that it does not reflect the variety of employment options in an area. A large and specialized manufacturing company may be located near a block group, which would lead to a large number of the employment access index. However, these may be highly specialized jobs, and the available employment nearby for the average worker may be very low. In order to improve this measure, it would be useful to derive a measure of employment diversity to interact with the measure of employment access. Zip code level county business patterns present one potential option.

However, as a simple measure of raw employment access, this index is well constructed.

2.3.6 NET RESIDENTIAL DENSITY

In addition to the gross household estimate described in Table 3, a more nuanced measure of residential density is estimated by CNT. Here blocks within a block group are denoted as residential blocks if they have at least one household per acre⁷. Net residential density is then measured as the average density just for those blocks within the block group that are residential. This measure is a worthwhile and well constructed improvement over the simple measure of residential density.

2.3.7 TRANSPORTATION COSTS

There are three necessary costs that must be estimated for the index: C_{AO} , C_{AU} , C_{TU} , which are auto ownership costs, auto use costs, and transit costs, respectively.

Auto ownership cost and auto use costs come from the 2007 AAA “Your Driving Cost” report. Ownership cost includes a variety of costs, including insurance, taxes, depreciation, and financing costs. Auto use cost is based on the cost per mile driven reported by AAA, with gas costs subtracted and regional gas costs added to allow for some regional variation in operating costs.

The lack of local variation in auto use and auto ownership costs (beyond gas price) is potentially problematic. For instance, missing from auto use cost is the time cost of a vehicle mile traveled. If one area suffers from more traffic congestion than another and this extends commuting time, then this could reflect

⁶ This distance was chosen as the 90th percentile of commuting distances from the LED dataset.

⁷ Different cutoffs for minimum density were tested by regressing the density measure corresponding to different cutoffs on auto ownership. The cutoffs ranged from more than zero households per acre (e.g., at least one household in the block) to 2 households per acre. CNT reports that the chosen cutoff was found to be “optimal”, but optimality was not defined.

an important cost difference between the two.

In addition there is parking cost, which may be both directly reflected in parking fees, and indirectly reflected in the time cost of parking as a function of parking supply. Given that the available supply of parking in a neighborhood is an important transit consideration to for many potential households, putting a dollar value on access both in neighborhood and in nearby destinations would be useful improvement.

Transit use costs come from the 2007 National Transit Database, which reports transportation revenues for specific transit agencies. If a transit agency serves multiple metro areas, the revenues were allocated to metros based on the distribution of bus and rail stations. Then the revenues were allocated to block groups within the metro areas based on the block group level percentage of transit commutes and household commuter counts from the ACS. The result is average household transit costs, which are then normalized by regional income measures as in the housing cost index.

2.4 H+T INDEX

The final H+T Index for each census block group is then the sum of the housing cost index for that block group and the transportation cost index for that census block group divided by the relevant income level.

3.0 ANALYSIS OF THE RIGOR OF THE H+T INDEX: MODELLING TRANSPORTATION USES

This section gives a broad perspective on what an index should be able to do and compares the H+T index to that ideal, and reviews the econometric underpinnings of the index, examines predictions of the model, and compares the model's results with transaction level real estate data for Philadelphia.

3.1 PERSPECTIVES ON THE TRANSPORTATION COMPONENT OF THE H+T INDEX

Creating an index of the spatial variation in transportation costs is fraught with challenges, primarily because any measurement of transportation cost at a particular geographic location is dependent, in part, on the destination and mode choices made by the residents of the location. Moreover, residents sort themselves into locations based on their preferences, destinations, and housing needs. These considerations raise a number of important questions that must be addressed in constructing an index of transportation costs that is intended for use in household decision making, underwriting or policy development.

An evaluation of a geographic index of transportation costs requires a clear understanding of what the characteristics of an ideal index would be, an assessment of the deviation of the actual index from the ideal, and an evaluation of the implications of that deviation for the uses of the index. In this section, we examine the conceptual challenges inherent in an effort to create a geographic index of transportation costs.

3.1.1 CHARACTERISTICS OF AN IDEAL GEOGRAPHIC TRANSPORTATION COST INDEX

Ideally, a geographic transportation cost index would:

- Reflect current per mile auto operating costs, auto ownership costs, and per mile transit fares;
- Adjust for the relative quality of the transportation experience;
- Include time cost of travel;
- Measure transportation costs for households choosing their optimal travel modes, destination choices, and private transportation capital investments;
- Accurately reflect changes in underlying transportation costs over time;
- Be independent of the composition of the households in a community.

Satisfying each of these criteria poses challenges, and collectively satisfying all criteria is very difficult.

Current Auto Operating Costs, Auto Ownership Costs, and Transit Fares

Measurement of the per mile travel costs by mode is the most straight forward input to the index. Average driving costs per mile, ownership costs and transit fares per mile are all available; and may not differ greatly from neighborhood to neighborhood. Unfortunately, even these basic costs are likely to be affected by household choices. Commuters that use a car to travel to a downtown job, for example may encounter

significant parking fees that raise the average per mile travel cost compared to a commuter traveling to a destination where parking is free. With the exception of parking costs, however, it should be possible satisfy this criteria.

Quality of Transportation

Not all travel is created equally. Some commute trips may entail greater disutility than others. For example, a transit trip with a seat is more desirable than one with standing room only. A trip that has lower out of pocket cost, or even lower time cost may not be as desirable as a more comfortable, more expensive trip. In general, efforts are made to control for quality in price indexes (as, for example, is the case with housing), and it is conceptually feasible in the context of travel as well, but the data are probably insufficient to do so. There is an additional complicating quality issue: consumers can invest in private transportation to make their travel more pleasurable. Thus quality is, in part, dependent on household choices. This complication, however, can be largely overcome by focusing on a standardized type of car (which is the primary type of private transport investment.)

Time Cost of Travel

The time cost of travel introduces serious complication into the idea of a geographic transportation cost index. Travel time cost is typically thought to be an increasing function of wages. This means that the travel time cost is unique to individual households, and thus depends on the households that choose to live in each neighborhood. This issue can be addressed, in part, by focusing on a representative household, and evaluating the geographic differences in travel costs for a representative household. However, this implies that relative transportation costs across neighborhoods may differ in their ranking may differ depending on the income the representative household.

Optimal Travel Mode, Private transportation Investment, and Destination

Households choose travel mode, transportation investments, and destinations to maximize their utility. These choices are a function of transportation cost, in part, but also factors specific to households. Metropolitan areas have multiple employment centers and travel modes and households living in the same proximate neighborhood will choose different destinations for employment and other travel. In addition, Households differ in:

- Structure: Families may have multiple workers and differing number of children;
- Preferences: People have preferences over public and private transport, auto quality, and destination choices;
- Income: Commute time may exceed out of pocket transportation costs and there is considerable variation in income within communities;
- Capital Constraints: Some households may be unable to choose their optimal transportation mode.

Heterogeneity among households implies that geographic cost rankings will differ across household type. Thus any single index may not be relevant to the marginal household considering whether to locate in a given community. For example, if everyone in a community took transit to downtown to work, travel costs would appear low in the community, and would, in fact be low for current residents. However, the travel

cost for a person requiring a car to get to their destination would experience different travel costs from those of existing residents. On the other hand, if the same community had a preponderance of people already driving, costs would be more representative of those of the prospective household. Creating an index that is generally relevant to all or most households is a fundamental challenge to constructing a transportation index. It should be noted as well that if an index is to be useful for either transportation investment decision making or for policy analysis, it should capture the transportation pricing incentives that households are confronting in their location and travel choices.

Changes in Neighborhood Travel Costs over Time

Conceptually, it should be possible to track changes in travel costs over time and to the extent that the index is driven by auto operating costs, auto ownership costs and transit costs, these data should be readily available. However, to the extent that these cost are affected by changing land use, new development or decline, changing economic structure, and changing infrastructure utilization levels, the data for the required adjustment are likely to be difficult to assemble and unavailable in a timely manner.

Independence from the Composition of Households in a Community

A geographic transportation cost index must control for differences in households across communities. Consider two neighborhoods, both equally well served by transit to a single downtown destination. One community has residents that only work downtown and commute by transit. The other community is also near a suburban center with only auto access. If one were to examine both communities, auto ownership would likely be lower where everyone commuted to the CBD by transit than in the community where some people also commute by car to the suburban location. But in this instance, we would observe greater car ownership costs in the community with suburban commuters and hence a higher transportation cost index, even though the commuting cost by transit is identical, and the community with suburban commuters has an additional commuting option. Thus the underlying observed data would be misleading with respect to commuting costs. It is crucial that cost index be independent of the household composition in a neighborhood in order to accurately reflect underlying travel costs.

The fact that different households in any neighborhood will have different transportation costs and the fact that locations evolve over time pose serious challenges for any transportation index that are difficult to overcome for several reasons. First, reliance on empirical observations of travel patterns in a neighborhood will necessarily capture the average or typical cost for households—but it will not necessarily reflect the cost of the marginal households considering moving to or from the neighborhood. Second, the household composition, employment destinations, and transportation costs associated with neighborhoods change over time, in part as a result of housing, transportation, land use and policies. Third, the households will sort themselves based, in part, on whether the location minimizes their transportation costs so that households in an existing neighborhood may not be similar to those considering moving to a neighborhood after a transportation investment or policy change.

Finally, it is crucial to recognize that transportation costs and housing costs are tightly linked. The fundamental insight of the canonical model of urban economics is that housing prices adjust to reflect differences in transportation costs. If housing markets are efficient, transportation advantages should be reflected in housing prices. The real world is substantially more complex than the monocentric urban model. There are multiple nodes for employment, shopping and recreation within a metropolitan area,

housing needs vary by type and size of household, available supply of housing is shaped by history and regulation, and tastes differ over modes of transportation. Still, one would be ill advised to ignore its basic insight when considering how to best construct an index of the housing and transportation cost vary spatially, and perhaps more importantly, to understand the general equilibrium consequences of investments and policies that change underlying travel costs.⁸

Because a household's transportation costs are a function of its optimal transportation-related choices, one could argue that separate indexes for discrete classes of households are required to satisfy the properties of the ideal transportation costs listed above. For example, one could estimate total transportation costs for a household with two adults and two cars, has one worker who drives to a downtown employment location, where he/she pays for parking, and earns \$30.00 per hour, and the adult drives 100 miles per week for other household purposes. This household would have markedly different costs than another in the same neighborhood that owns one car, takes transit to a downtown location, earns \$30.00 per hour and travels 100 miles per week for household purposes, even though the households have the same origin and destination. Differences across households are likely to be even greater when differences in destinations, household structure, preferences, and income and capital constraints are considered. Obviously, there are a very large number of potential permutations of potential transportation costs for people living in a given neighborhood.

3.1.2 DEVIATION FROM THE IDEAL INDEX

CNT's Approach to Constructing a Transportation Cost Index

The CNT H+T Affordability Index seeks to address many of the criteria for an ideal transportation index while providing a single index of transportation costs. Specifically, the CNT index reflects current estimates of auto operating and ownership costs and transit costs, addresses differences in households' characteristics (including time costs through controls for income), and reflects the optimal travel choices of the representative household. The index does not seek to control for differences in travel quality or changes in transportation cost over time.

CNT's approach to the transportation component of the H+T index is regression based. It is based on three separate models of household transportation choices:

- Vehicle miles traveled
- Number of autos per household

⁸ The prediction of the "monocentric model" is more than an academic curiosity. Consider for example the interaction of housing, transportation and traditional mortgage underwriting standards in the context of the basic insight of the urban model. One can choose to live further away and pay a lower price for an otherwise similar house closer to the center. "Drive until you qualify" results in a lower required down payment and has no impact on the income requirements of lenders. However, given current underwriting standards the effective income constraint is less restrictive for more distant houses because income after transportation costs is lower than for those of similar incomes choosing houses in closer proximity to the center. Thus the monocentric model itself points up the need for a broader set of underwriting standard that are inclusive of transportation costs.

- Transit share of travel

Each dependent variable is modeled as a function of neighborhood characteristics, which include measures transit access, density, and employment as well as other features. In addition the models include household characteristics such as income and other demographic variables. Inclusion of household characteristics allows the estimation of expenditures for a “representative” household, which, as discussed above, is crucial to creating an index that is independent of the composition of households in a neighborhood. The econometric detail and performance of these models are discussed in a subsequent section of this report.

Aside from technical issues of model specification and data quality, there several features of the regression approach that need to be carefully considered. These include:

- Usefulness of a single index
- Measures of the accuracy of the model
- Understanding the implication of unmeasured household and transportation factors
- Capitalization of transportation characteristics into property values.

Usefulness of a Single Index

The CNT H+T index is based on the transportation choices of a representative household in each neighborhood as defined by a census block group. By modeling transportation as a function of household, transportation, and neighborhood characteristics, CNT can construct an estimate of the costs associated with the representative household, based on the characteristics of the transportation system and neighborhood. This approach has an appeal in that it yields a single estimated cost for each geographic location that can be readily compared across neighborhoods.

To the extent that transportation costs for the representative household are strongly correlated with the costs of households considering locating in the community, the index is potentially useful for households comparing neighborhoods. However, transportation costs are likely to vary substantially from household to household, and it is conceivable that the relative ranking of transportation costs across neighborhoods could vary substantially as well. The regression framework does provide a potential solution to this concern. The regression model could be used to calculate the expected transportation cost in a given neighborhood for households with any combination of characteristics. This concept, which is readily implementable, could address concerns regarding the appropriateness of the index for individual households.

Measures of the Accuracy of the Index

Evaluating the accuracy of a transportation cost index that attempts, appropriately, to control for the composition of households in a neighborhood is inherently challenging. Standard economic methods are useful for evaluating the statistical properties of the model, but do not provide much insight on the extent to which the regression models result in reasonable transportation cost indexes, even for the representative households. Comparing indexes of transportation costs with neighborhood averages is clearly not enlightening since, by construction, the index seeks to evaluate the cost of a representative household

rather than the costs for the mix of households in the neighborhood. One approach to addressing this issue is to use the regression model to predict what transportation costs are for the actual population and compare the predictions with the actual. This provides a measure of the magnitude of potential errors in the index. Ultimately, the believability of the index rests on how well it predicts transportation costs for actual households. We examine this question in the technical discussion below.

Understanding the Implications of Unmeasured Transportation and Household Factors

Prediction errors raise other issues. If prediction errors are systematic, then it implies that something is missing from the statistical model that could impart systematic bias to the transportation cost index. For example, transit quality may not be adequately captured in the transit model, with the likely occurrence that the predicted transit use in area with high amenity (unobserved) transit service will be underestimated, and therefore transit costs overestimated for the community. Note that this is a problem that is a result of the effort to control for differences in household characteristics; otherwise one could just use observed transit ridership, which would reflect the higher quality service. Thus the effort to control for household differences across neighborhoods is not without cost. Patterns in prediction errors are evaluated in subsequent sections.

Capitalization of Transportation Characteristics into Property Values

Economic theory and empirical evidence shows that local advantages in transportation result in higher property values. Thus transportation advantages are likely to be offset by higher house prices. Because the H+T index is based on housing and transportation cost, the index reflects total costs, including any offset of transportation through property values. However, the relationship between house prices and transportation costs raises questions about the likely impact of policies and transportation investments designed to lower transportation costs on the total affordability of both housing and transportation. In particular, changes in investment or policy that reduces transportation costs will be reflected in the transportation cost index. However, there would be no corresponding change directly implied in housing costs, since this component of the index is not regression model based.

3.2 REGRESSION FRAMEWORK FOR TRANSPORTATION MODELS

The transportation model utilizes a non-linear regression to estimate how various environmental and household characteristics explain the three measures of transit utilization. In order for these models to generate the correct predictions, they must accurately characterize the relationship between the independent and dependent variables. In order to determine whether this is the case, there are several econometric issues which must be addressed.

The basic form of each of the utilization models is a rational function regression, which models the utilization as the following:

$$F_z = \alpha_0 + \sum_{j=1}^m \frac{\alpha_{1j} x_j + \alpha_{3j} x_j^2 + \alpha_{5j} x_j^3}{1 + \alpha_{2j} x_j + \alpha_{4j} x_j^2 + \alpha_{6j} x_j^3} + \xi$$

Where x_1, x_2, \dots, x_m are the independent variables in Table 3 and z indexes AO , AU , and TU , the three utilization categories. There are several issues with this specification that will be discussed below.

3.2.1 MODEL FORM

In regression analysis economists usually begin with the presumption of using ordinary least squares (OLS). This statistical technique is the workhorse of economics, in part because it can be justified as optimal based on a wide variety of assumptions and criteria, and in part because it has many desirable statistical characteristics. For instance, given some basic assumptions about the data⁹, OLS is unbiased, and the most efficient estimator. With even fewer assumptions, least squares can be shown to have very desirable asymptotic properties that hold when sample sizes are sufficiently large.

In addition to the desired properties of OLS, the model is also the linear expression of the conditional expectation of the dependent variables. This means that the coefficients of an OLS estimate are easy to interpret. This is true even when independent variables enter non-linearly in the regression (e.g. when using squared or log transformations of independent variables), as the linearity of the model allows for easily estimable and understandable first derivatives that show a change in the independent variable changes the expected value of the dependent variable.

OLS also has the advantage of a closed form unique solution. In comparison, other models, including some nonlinear models, are often solved using iterative maximization processes, which may not have a unique solution. Thus statistical packages using different maximization routines may arrive at different conclusions.

Because of these desirable and well understood properties, applied economics work typically begins with the presumption of OLS and moves from there to other methods based on specific arguments for why the alternative model is a better fit for the particular data or purposes. This means that one should provide reasonable justifications for departing from OLS showing that a different modeling technique provides reasonable benefits over OLS. Often this takes the form of arguing that OLS is inefficient, or biased based on the failure of one or more assumptions. In these cases alternative model can sometimes be found that under these circumstances maintains some desirable properties, or is known to be more desirable (e.g. more efficient, less biased), than OLS.

The main advantage of the rational functional model used by CNT is that it is extremely flexible and can approximate many complex functions. However, one can retain the desirable properties of OLS and still produce highly nonlinear and flexible relationships. For instance, an alternative to the above model is to use the form:

⁹ This includes normality of the error terms, linearity of the data generating process, uncorrelated errors, and exogenous independent variables that are not perfectly collinear.

$$F_z = \beta_0 + \sum_{j=1}^m \beta_{1j} x_j + \beta_{2j} x_j^2 + \beta_{3j} x_j^3 + \beta_{4j} x_j^4 + \eta$$

This allows a complex and nonlinear relationship between independent and dependent variables (but not between coefficients, which are strictly linear) while retaining the desirable properties of OLS.

In addition, the model does not provide significantly different results from the rational functional form model. For the autos per household regression using transit variables, the R-squared for OLS is 0.73 and the R-squared for the nonlinear model is 0.74. The predicted values resulting from these models are also extremely close, which implies the impacts on the index values from using OLS will not be significant. In fact, over 90% of the OLS predictions are within 0.08 autos per household of the nonlinear model predictions. Given the average auto per household of 1.7, this difference is relatively minor.

In addition, the correlation between the predicted values resulted from these two regressions is nearly perfect at 0.99. This is important because the index values are more useful as relative comparisons than absolute values, and the high correlation suggest that relative differences between locations are likely to be preserved.

While there is no specific decision criteria to decide whether the increase in R-squared justifies using the nonlinear model, it is our consideration that the small increase in explanatory power does not justify the choice of a nonlinear rational function model over OLS. This is especially true given the lack of consequential differences in predictions and thus index values.

3.2.2 NON-RANDOM ERRORS

Another issue with the model used in CNT's regression is that the errors are not likely to be fully independent from each other. Within a given city, for instance, if auto use is higher than the model predicts in one block group then it is likely to be higher than predicted in other, nearby block groups. This can result in a variety of statistical issues, and suggest that different modeling approaches may be desirable.

One common concern is that errors which are correlated among groups, for instance within cities, will lead to incorrect standard errors. Error correlation should not bias the coefficients and thus predictions per se, but will lead to incorrect standard error measures. In the end this could quite possibly lead to a sub-optimally specified model as standard errors are used for CNT's iterative process of selecting which variables to include in the regression.

In these circumstances, one approach is use clustered error variance estimates. This adjustment can be done using a geographically defined group variable, for instance all block groups in a county.

For instance, using the linear regression model defined above, the following variables are all statistically significant and so should be included in a final regression if one were to use statistical significance to determine whether variables are included as CNT does:

Table 6: Coefficients Made Insignificant From Clustered Errors

Variable	Exponent
Average household size	3
Employment gravity index	4
Intersection density	1
Transit access shed	1,2,3,4

However, using county groupings to create clustered standard errors, each of these becomes insignificant and would not be included in the final regression using these criteria¹⁰.

Another common concern is that there will be some group level unobservable characteristic that is correlated with the independent variables. A plausible candidate in the transit utilization regressions would be citywide unobservable differences that are correlated with either household or transit independent variables. In the linear regression framework, this would take the form of:

$$F_{ij} = f(H_{ij}, \Psi) + g(E_{ij}, \Pi) + \eta_{ij} + \nu_j$$

Where i indexes block groups within j cities, and H_{ij} is a vector of household specific variables. If ν_j is correlated with any variables H then the estimates of W will be inconsistent, and if u_j is correlated with E_{ij} then estimates of B will be inconsistent. In this case, a fixed effects estimator could be used to obtain consistent estimates of the coefficients.

If group level unobservable characteristics exist but are not correlated with independent variables, then a possible alternative to fixed-effects estimation is random-effects. This allows for group level variation in the error term, but does not require the estimation of group level constants, making the model more parsimonious.

Whichever model is used, it should not be presumed that OLS is the preferable model compared to fixed effects or random effects when there is a plausible likelihood of group level errors. Instead this question should be determined using the appropriate statistical tests.¹¹ However, it may be the case that changing the model will not significantly impact the results. Predictions using a fixed effects regression and the OLS model discussed have a correlation of 0.98, suggesting that changing to fixed effects by itself will not greatly alter the results.

Rather than discrete group based correlated errors, it may be more accurate to model the errors as being spatially auto correlated. Examining maps of the residuals from the auto ownership regression, it appears likely that areas suffer from some level of non-random spatial patterns.

¹⁰ Using the specific process that CNT used this is not necessarily the list of insignificant variables that would end up included in the final regression, however it is illustrative of the problem of failing to cluster and statistical significance, especially when this is a variable selection criteria.

¹¹ Statistical tests, for example the Hausmann test, can help determine if the fixed effect or random effect model is sufficient. It is possible, although seemingly unlikely, that OLS would be sufficient, which can be determined relative to the fixed-effects regression using F-tests, and relative to the random effects model using the Bruesch-Pagan test.

Figure 2: Auto Ownership Residuals: Philadelphia

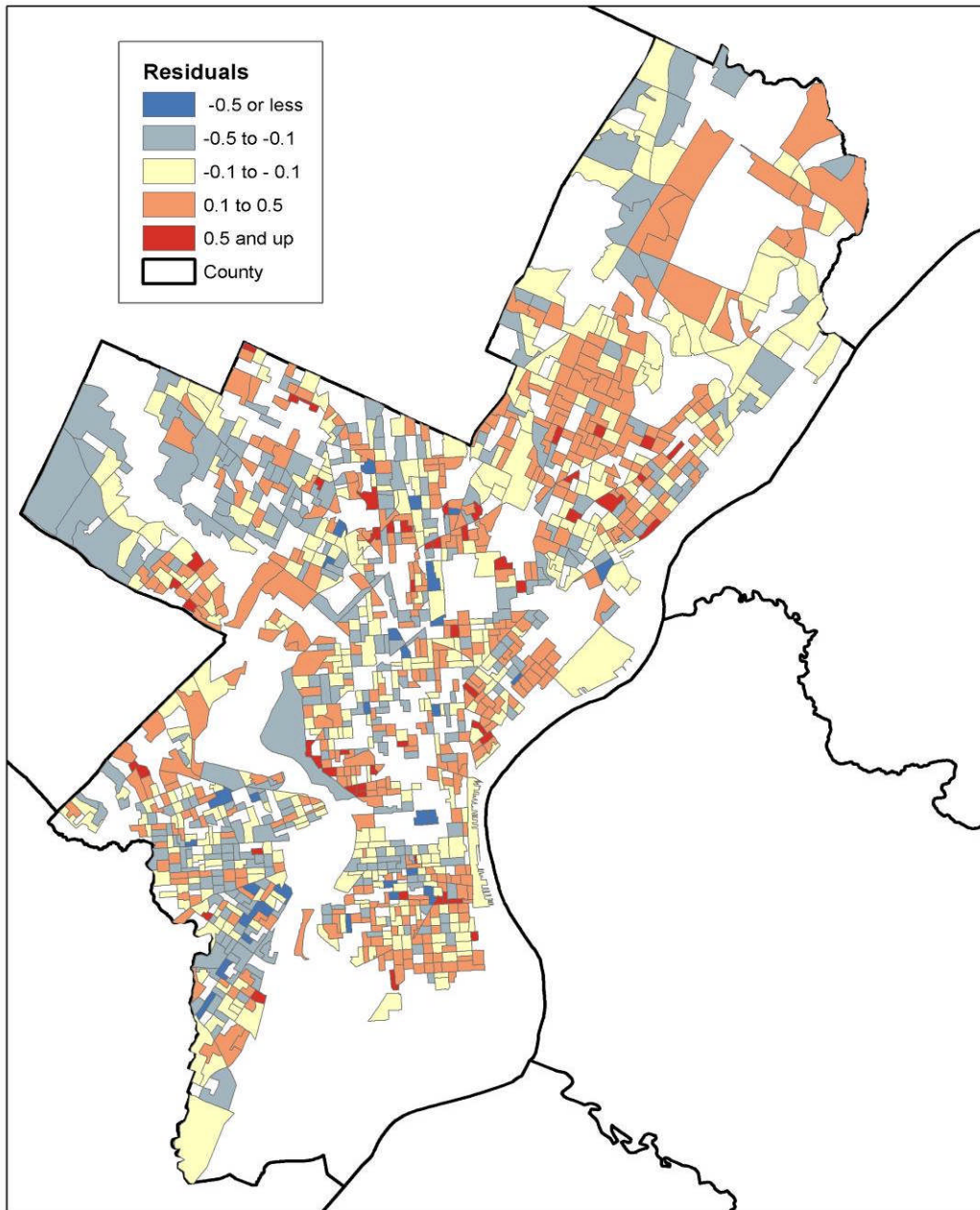


Figure 3: Auto Ownership Residuals: Seattle

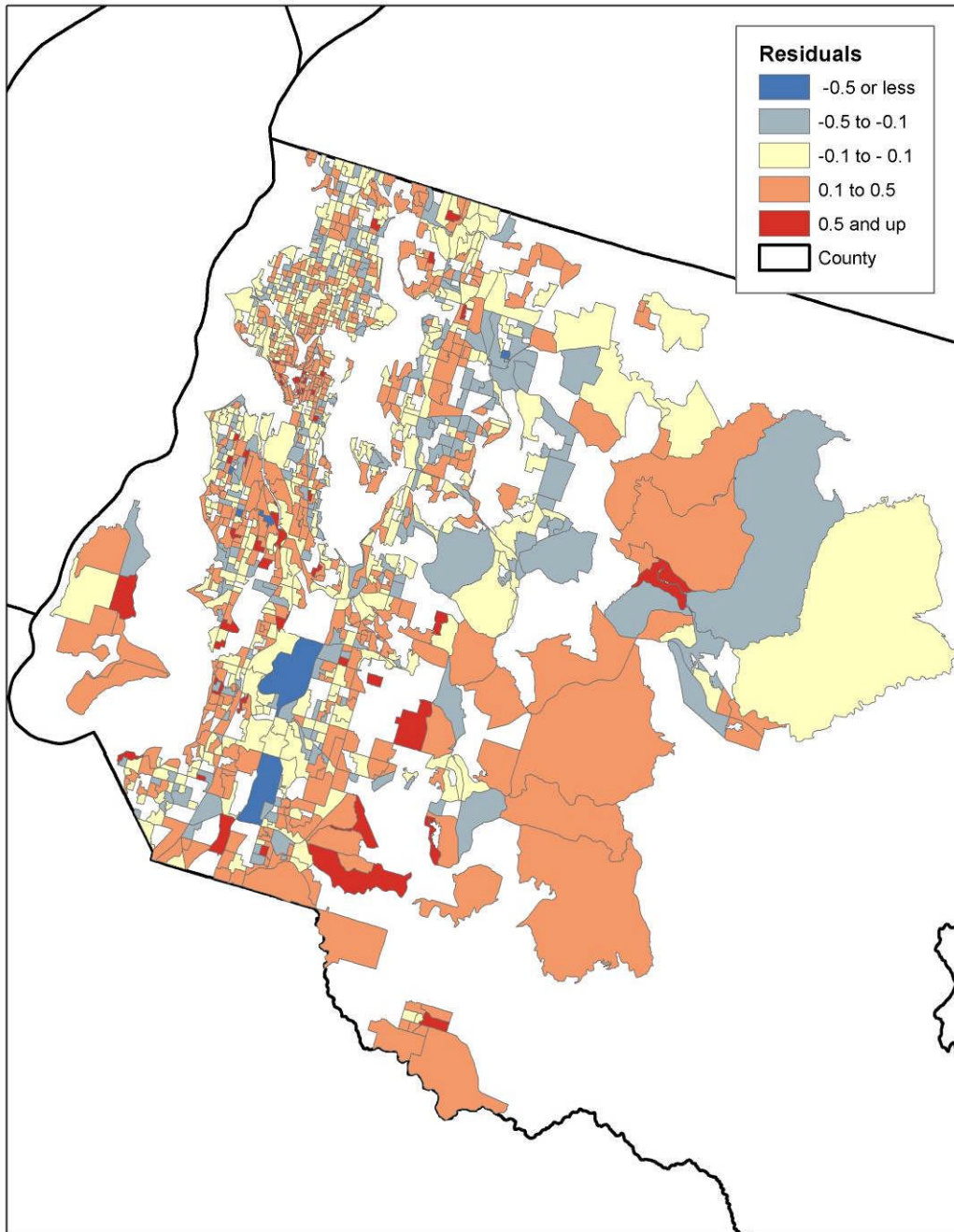


Figure 4: Auto Ownership Residuals: San Diego

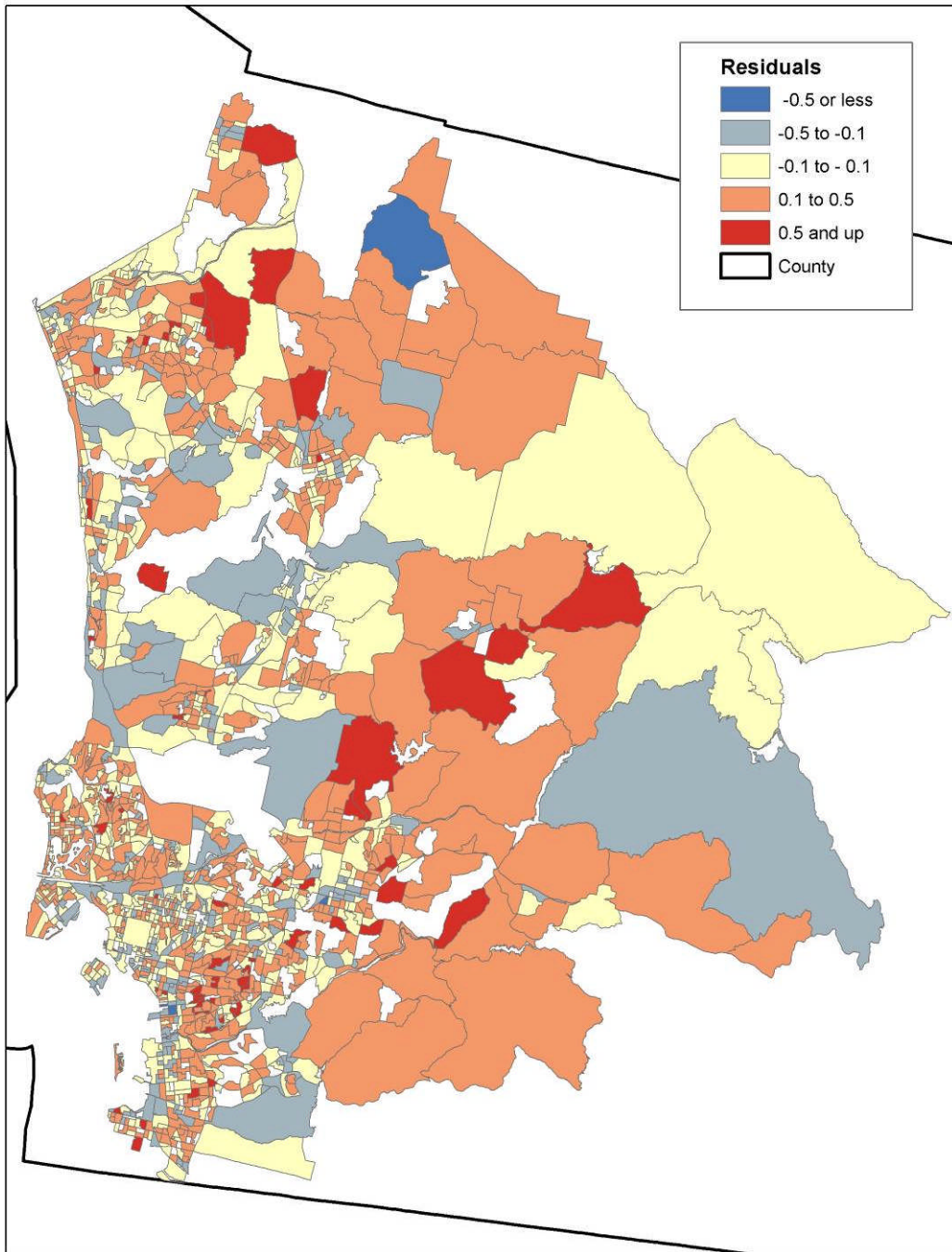
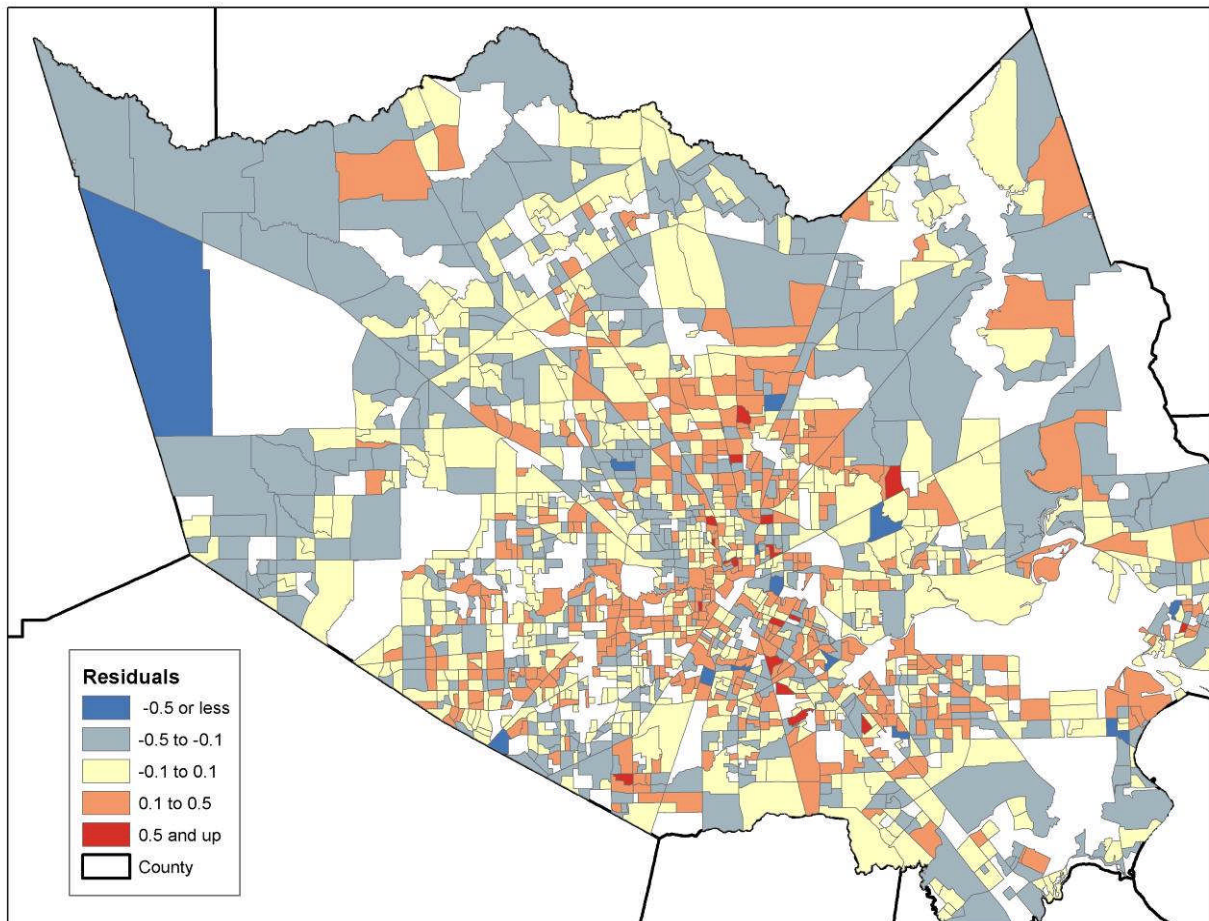


Figure 5: Auto Ownership Residuals: Houston



While there clusters do appear, it is also clear that there is a large degree of randomness within the city, which indicates that the variety of spatial measures in the model have helped control for much of the spatial correlations. Moran's I test for spatial autocorrelation was run on the residuals for auto ownership for a sample of cities, with the results presented in Table 7 below. The results indicate that there is a statistically significant degree of spatial autocorrelation in the data in each location. However, this does not mean that the controlling for the spatial correlation would result in economically significant different results. Nevertheless, it suggests that models controlling for spatial autocorrelation should be explored for the CNT index.

Table 7: Spatial Autocorrelation for Selected Cities

City	Moran's Index	Z Score	p-value
Philadelphia	0.136	18.611	0.00
Houston	0.0651	17.472	0.00
San Diego	0.042	12.71	0.00
Seattle	0.069	18.81	0.00
New York	0.051	117.87	0.00

Note: For all but New York, city is defined as primary county of the city.
For New York the Core Based Statistical Area (CBSA) was used, which includes areas of northern New Jersey.

The discussion and evidence in this section suggests that there may be geographical correlation and group clustering issues that should be examined. However, the residual maps suggest that whatever correlations exist, there is a large degree of randomness in the residuals. In addition, the use of alternative regression models, including OLS, fixed effects, and random effects, should be explored. However, the high degree of correlation between predictions generated from alternative models and the CNT models suggest that in the end correcting for these differences may not have significant effects on the index results. Nevertheless, these are issues that deserve some exploration and merit further research.

3.2.3 OTHER MODELING ISSUES

There are two additional concerns with the transportation model. The first concerns the lack of region specific variables for the transportation estimator, and the second covers the similarity of the test markets for the VMT model.

The coefficients of the transportation model are not permitted to vary by region, and independent variables do not interact. It may be the case, for instance, that residential density has a different impact on auto ownership in different regions of the country, or that the impact of employment density varies with residential density. The current model specification will not recognize these differences. If there are regional differences, the model fit could potentially be improved by exploring modeling options that included region specific variables.

Additionally, the VMT model coefficients were estimated using data from Massachusetts. The estimated coefficients on environmental and household variables are then applied to the independent variables for all other locations to generate predicted values. CNT reports that they tested these estimation results against those generated similarly using vehicle miles traveled data from Chicago and found they generated similar predicted values. While this test does provide some evidence of the robustness of the model, Chicago and Boston are fairly similar to each other, as they both have extensive transit infrastructure and a wide variety of densities. However, Massachusetts and Chicago are very different from many other areas in the country. It is unclear whether the results would stand if data from a Southern city, or suburban or rural area was used. Whether the coefficients from Boston are useful for the rest of the country should be further tested to the extent data will allow in other, more different regions of the country.

3.3 PREDICTIONS OF THE TRANSPORTATION COMPONENT

For each category of transit utilization we can examine the transit utilization that is not explained by the regression model. First, define each of the three utilization variables as:

$$F_z = f(H, \Psi) + g(E, \Pi) + \eta \quad (1)$$

Where f is a non-linear function of the vector of household demographics H , and the coefficient vector Ψ . Likewise, g is a non-linear function of the vector of environmental variables, and Π is the coefficient vector. The unexplained portion of utilization is contained in the residual η .

The predicted level of utilization is based on observables is then:

$$\hat{F}_z = f(H, \hat{\Psi}) + g(E, \hat{\Pi}) \quad (2)$$

Where H and E are the actual local demographics and environmental variables, and $\hat{\Pi}$ and $\hat{\Psi}$ are the estimated coefficients. The index, in contrast, is based on:

$$\tilde{F}_z = f(\tilde{H}, \hat{\Psi}) + g(E, \hat{\Pi}) \quad (3)$$

Where \tilde{H} is a vector of regional average demographics. However, the η contains information about unexplained differences in transit utilization, some or all of which may reflect positive transit performance that is not captured in the environmental variables and as such should be included in the index. The only reason for excluding η is if it is a function of unmeasured demographic variables which should then be controlled for. If instead this reflects unmeasured positive aspects of the transit infrastructure or other neighborhood characteristics that lead to different amounts of auto ownership or transit utilization, then it should be included in the index.

Based on the R-squared reported by CNT, these residuals will be around 25%, 40%, and 15% of the variation for autos per household, percent transit commuters, and vehicle miles traveled¹². It is straightforward to estimate utilization numbers that reflect these unobservables by adding the residual from (1) to the predicted value from (3):

$$F'_z = \tilde{F}_z + \hat{\eta} = \tilde{F}_z + F_z - \hat{F}_z$$

This new measure has only a 0.73 correlation with the existing measure for \tilde{F}_z for auto ownership, and a 0.80 correlation for percent of households commuting. This suggests that the inclusion of the entire residual in the index measure could modestly change the results of the index.

¹² Note that for this last category residuals are only available for Massachusetts, so this does not in general apply to this category.

Or one could assume that some proportion δ of the unobservable utilization is due to unmeasured environmental factors and some percent is due to unmeasured household demographics or other variables that should be excluded from the index. In this case the index is estimated as:

$$F_z'' = \tilde{F}_z + \delta \hat{\eta} = \tilde{F}_z + \delta (F_z - \hat{F}_z)$$

Using a value of 50% for δ , these correlations are 0.91 and 0.93 respectively. This translates to a change of 0.15 autos per household or more for 20% of the observations. The percent of households commuting changes by 3 percentage points or more for over 20% of households. As expected, using a value of less than 100% for δ results in less substantial but still non-trivial changes in the utilization estimates.

It does not seem likely that the entire amount of the unexplained variation in transit utilization is due to household demographics and other variables that should be controlled for. Therefore some number for δ greater than zero should be used¹³. One possible way to estimate this is to examine the percent of the explained variation in utilization, that is \hat{F}_z , is explained by demographic variables and what percent are explained by environmental factors, and then assume a similar proportion holds for the unexplained variation.

In the previous section, a fixed effects estimator was proposed as a possible regression model. This takes the form:

$$F_{ij} = f(H_{ij}, \Psi) + g(Ei_j, \Pi) + \eta_{ij} + \nu_j$$

Where i indexes block group, and j indexes city or some other geographic area. One concern with a fixed-effects estimator might be that average unexplained levels of utilization in each city, $\hat{\nu}_j$, is included in the predicted utilization levels instead of in the residuals, as would be the case with OLS. This returns to the issue addressed above about whether unexplained variation in the utilization measures should be included in the index or removed by keeping the residuals out of the prediction. However, the current CNT model includes a constant, and the fixed effects estimator simply allows the constant to differ by geographic area. Thus given the inclusion of the national level fixed effect in the index, e.g. the constant, it would be consistent with their current measure to include the city level fixed effect. Again, one could use a fraction of the fixed effect estimate to allow some portion to remain in the residual.

3.4 CASE STUDY: PHILADELPHIA DATA

Using transactions data from Philadelphia, the CNT housing cost measure can be compared to both local median transaction prices and local house price indices measuring appreciation.

¹³ A value of zero for δ is equivalent to using the current CNT measurement.

If appreciation rates vary geographically, then the relative housing and transportation affordability of neighborhoods would change over time, suggesting that the static and lagged nature of the index may be problematic. The transactions data was used to estimate tract level constant quality price indexes, which were then used to estimate appreciation rates to 2011 from every year from 2005 to 2010. In order to have a similar geographic level of comparison, the housing cost measure from CNT is aggregated to the tract level using the number of occupied housing as a weight. As shown in Table 8 below, appreciation rates are unrelated to housing cost measures for the 291 Philadelphia tracts where price indexes could be estimated:

**Table 8: HPI Appreciation Correlation
With Housing Cost Measure**

Change	Correlation
2005 to 2011	-0.08
2006 to 2011	-0.10
2007 to 2011	-0.07
2008 to 2011	-0.04
2009 to 2011	-0.06
2010 to 2011	0.03

The results suggest that the indexes measure of housing cost is not related to appreciation rates, and therefore do not suggest that the relative affordability rankings systematically changes, at least over the time horizon that we investigated. These results are confirmed by median transaction prices at the tract level. Of course these preliminary findings cannot be generalized beyond Philadelphia without additional research.

In addition, the transactions data can be used to test whether the CNT housing cost measure is related to housing price levels. Since housing transaction costs will be strongly related to down payment and mortgage payments for new homeowners, it is desirable that this correlation be strong and positive. As the Table 8 below indicates, housing costs are in fact strongly positively correlated with median transaction prices in every year.

**Table 9: Median Transaction Price Correlation
With Housing Cost Measure**

Change	Correlation
2005	0.81
2006	0.81
2007	0.82
2008	0.72
2009	0.80
2010	0.82
2011	0.80

In addition, despite the volatility of median house prices (standard deviation of percent changes from 2010 to 2011 is 0.42%) the correlation is relatively stable. However, it should be noted that there is a significant difference between CNT's housing cost measure and the median transactions price. A regression of 2011 median prices on the housing cost index has an R-squared of 0.60, meaning 40% of the variation in house prices cannot be explained by median housing cost index.

4.0 EXISTING HOUSING AND TRANSPORTATION INDEXES

Housing affordability indexes and calculators are widely available and used to track the affordability of housing across regions and over time. Here we review the major housing affordability indexes and examine what they do and do not measure. While several housing affordability indexes are in widespread use there is much less availability and use of alternative transportation affordability indexes. There are two which are not as widely used which we describe below. As discussed above by combining a measure of housing costs with transportation costs at the neighborhood (block group), the goal of the H+T is to fill a void by creating a tool to assess the affordability of a neighborhood based on these two largest household expenditures, to inform household decisions and public investment.. We describe alternative measures of affordability in 4.1 and provide a description of the principle alternative affordability indexes in section 4.2. We then turn to a summary of what we learned from a survey of users of the H+T index in 4.3.

4.1 ANALYSIS OF APPROACHES TO INDEXES

Housing affordability indexes have been developed by a number of professional organizations to monitor the level and evolution of housing affordability defined as a household's ability to qualify for mortgage financing. Among the established indexes are the Housing Affordability Index (HAI) developed by the National Association of Realtors (NAR) that measures if the median-priced home is affordable to a median-income household and the Housing Opportunity Index (HOI) developed by the National Association of Home Builders that calculates the share of homes in an MSA affordable to a median-income family. Alternative measures of housing affordability were subsequently developed (MIT's Affordable Housing Index, CNN's Housing Affordability Calculator). There are also tools that have been developed to reveal transportation costs in different locations such as the Walk Scores and the University of Minnesota Transportation Costs Measure.

The innovation of H+T in this regard is to add transportation costs to a measure of housing affordability. That is transportation costs also matter in measuring affordability and it is the combination of the two that impacts what is left in the budget for the purchase of other goods and services. The combination of the two also determines whether a mortgage can easily be paid off. This insight is behind the concept of location efficient mortgage as well (Danielson, 1999). Households do in fact drive to qualify, that is they avoid expensive mortgages by commuting longer distances, but that commuting cost is not captured in the way mortgage costs are quantified and related to income. The efficient mortgage approach attempts to do so, but on its own it does not provide market-wide information on affordability when both transportation and housing costs are taken into consideration. The contribution of H+T is to set up a framework to combine both measures in a single index.

Historically and for cross-country comparisons the most frequently used indicator of housing affordability is the simple ratio of house price to income based on the average house prices and income in a specific country. For the U.S., the National Association of Realtors introduced in 1970 a measure of housing affordability that augments this simple ratio by adding to it a mortgage borrowing cost component as described in 4.2.1.

The academic literature on housing affordability focuses on tenure choice and the extent to which constraints, as measured by affordability, impact tenure choice along with other economic variables such

as the cost of housing and household income and preferences. These studies generally show that the decision to become a homeowner is impacted by “constraints.” That is, the household may prefer to be an owner as opposed to a renter, but that household cannot “afford” to be an owner, either because the household lacks sufficient saving for a down payment or sufficient income for a mortgage payment for their “optimal” home. Thus the literature identifies whether potential homeowners are or are not income constrained, wealth constrained or, in the recent literature, constrained by their poor credit history. The NAR measure of affordability is an indicator of the degree to which borrowers in a specific market are likely to be income constrained and unable to afford to own a home. The advantage of this index over the H+T index is that the NAR index is current, updated with annual data, and relevant for potential homebuyers in that neighborhood, unlike the H+T index which is reliant on historical data for existing homeowners in that neighborhood.

Besides the NAR index, another measure of housing affordability that is widely used is the NAHB-Wells Fargo Housing Opportunity Index described in 4.2.2 below. Although these two indexes vary, they both point to where housing prices relative to income make homeownership less affordable to the median income household in a region. Housing prices vary greatly across regions in the US, within regions and even within localities. Both indexes are developed for regions and for the nation as a whole and not for municipalities nor for neighborhoods within municipalities. This is in part due to lack of timely data for house prices on the neighborhood scale. Metropolitan areas are also used for these indexes because housing markets are traditionally identified as being the entire metropolitan area, since homebuyers can in theory shop for housing across a metropolitan area which is defined as the commuting shed of a CBD. Thus the literature implies that there is choice to locate across an MSA when looking for housing, and that households are not constrained to a particular neighborhood within an MSA. The H+T index is created for the neighborhood scale as it should be if transportation costs are to be part of the location decision. Thus the H+T index must be on the neighborhood scale, unlike the NAR and HOI indexes. Nonetheless the currency of the data and the conceptual frame, based on a representative household and not the existing households, used by these indexes may be useful for strengthening the housing affordability component of the H+T index. Suggestions for doing so are included in recommendations.

A separate concept used to evaluate housing choices is rental or user costs for homeownership. Where and when user costs are high relative to rental costs, according to this approach, the marginal home seeker should choose to be a renter (Diaz and Luengo-Prado, forthcoming). Another strand of this literature expands on the concept of user cost to include public amenities such as jobs, public school, public safety to point to where a homebuyer (or a renter) gets the most for their money (Fisher et al, 2009). From this perspective households first choose a community and then a quantity of housing in order to maximize their utility over a bundle of goods and services such as public goods, commutes and housing prices (Banzhaf and Farooque, 2012).

The index described below in 4.2.3 is designed to provide a measure of the cost efficiency of housing (rental or owner occupied) across neighborhoods. Transportation and commuting costs are included in the index’s measure of neighborhood affordability. Because this and other public good amenities are included, this measure conceptually could be used to consider from a broader societal perspective the implications of location choices for policy relevant outcomes such as sustainability. This is particularly important for public transit investments and sustainable city policies such as HUD’s Sustainable Communities Regional Planning Grants program. While this index is conceptually useful it does not measure affordability

constraints as both the NAR and NAHB index do. Moreover, implementation of such an index on an annual basis for the nation as a whole would be a daunting task.

All indexes have their strengths and weaknesses in relation to three broadly defined measures: 1) Indicators of the degree to which homeownership choice is constrained; 2) indicators of where housing is cost efficient from an individual perspective; and, 3) indicators of where public investment should be put into place from a societal perspective, for example, to encourage settlement and denser communities. The goal of the H+T index is to be of use in each of these dimensions. Below we review existing alternative housing affordability indexes and calculators; we also note two other tools that measure transportation access and analyze how H+T compares on these three dimensions.

4.2 ALTERNATIVE MEASURES OF HOUSING AND TRANSPORTATION AFFORDABILITY

4.2.1 NAR HOUSING AFFORDABILITY INDEX

The National Association of Realtor's Housing Affordability Index (HAI) measures whether or not a typical family could qualify for a mortgage loan on a typical home in the nation or a region. A typical home is defined as an existing single-family home at the monthly median price, as calculated by NAR using a sample of Multiple Listing Service (MLS) sales for existing single-family homes. The typical family is defined as one earning the national median family income based on monthly income estimates from the census. The prevailing mortgage interest rate is the effective rate on loans closed on existing homes as reported by the Federal Housing Finance Board. In addition to a composite index, one based on fixed rate mortgages and another based on adjustable rate mortgages are also calculated.

These components are used to determine if the median income family can qualify for a mortgage on a typical home assuming a 20 percent down payment and that in order to qualify, the maximum mortgage payment (principal plus interest) must be 25 percent or less of the household gross monthly income. An index value of 100 means that a family earning the national median income can qualify for a mortgage on a median-priced home, assuming a 20 percent down payment. An index above 100 indicates that a family earning the median income can qualify for a loan with a price greater than the median. The formulas used to calculate NAR's HAI summarized here are available on their website.

The index's values are publically available going back to 2008, but the index has been calculated going back to 1970 and is updated monthly. Both national and regional (Northeastern, Midwestern, Southern and Western) indexes are calculated.

4.2.2 NAHB-WELLS FARGO HOUSING OPPORTUNITY INDEX

The Housing Opportunity Index (HOI) measures the share of homes in an MSA that is affordable to a family earning the MSA median-income. The HOI was developed in the early 1990s by the National Association of Home Builders in partnership with Wells Fargo. It is defined as the percentage of homes sold in a given area that is affordable to a family at the local median income level based on mortgage underwriting criteria

described below. In the HOI, housing affordability is defined using more region-specific information than NAR's HAI.

Income data come from the annual median family income estimates for metropolitan areas published by the Department of Housing and Urban Development. The assumption used is that a family can afford to spend 28 percent of its gross income on housing. Housing price information is obtained using sales transaction records from CoreLogic. The monthly payment (principal plus interest) that an owner would pay assumes a 30 year fixed rate mortgage, with a 10 percent down payment. The effective interest rate used is based on data reported by the Federal Housing Finance Board. The estimated monthly payment also includes an estimate of the property taxes and property insurance for a given home based on Census data on tax and insurance rates at the metropolitan level; however, the estimate does not include private mortgage insurance. For each property sold during the last quarter NAHB calculates an estimated monthly housing cost and the income required to afford it. The HOI is "the share of records in a metropolitan area for which the monthly income available for housing is at or above the monthly cost for that unit."

The index is published quarterly for the nation and for between 150-225 metropolitan areas going back to 1991.

4.2.3 MIT HAI AFFORDABLE HOUSING INDEX

MIT's Housing Affordability Initiative (HAI) measures the ratio of affordable units to the number of units in a town, which represents the town's affordability index. HAI affordable housing index differs from other affordability indexes in that it takes into account locational amenities in its pricing model. The authors of the methodology explicitly define the goal of affordable housing as not only to provide affordable structures but also units accessible to jobs, safe communities and good schools (Fisher et al., 2009), each dimension having its own implications. In this model the measure of housing affordability controls for the structural quality of homes, the needs of varying households as well as the commuting costs involved with a unit's location. The HAI affordable housing index considers the "total cost of ownership of individual housing choices. As a result, opportunity costs of access to employment, quality of public schools, environmental conditions, structural quality of units and the needs of households are factored into the index" (Fisher et al., 2009). The methodology developed to calculate this index takes into consideration the fact that housing prices and therefore affordability reflects commuter costs. This represents a conceptual advantage of this index.

The index was calculated for several New England metropolitan areas at the place level. The majority of the data is taken from the 2000 decennial census and is coupled with updated stock information on unit structural characteristics and tenure from The Warren Group that collects the data from city and town assessors. Information about commuting times and costs are obtained from several state agencies. House values from the 2000 census are updated using a combination of data from Case-Shiller repeat sales indexes at the ZIP level, Zillow.com estimates at the town level and a hedonic regression based on sales data collected by The Warren Group. Imputed rents are then calculated for owner occupied housing based on estimated current house values, interest rates and tax liabilities. Imputed rents are then used to compute the marginal values of town-level amenities such as job accessibility school quality, and open spaces.

The model imposes additional costs to homes that exhibit inadequacy in various fields and these unit-specific figures are then aggregated to produce town-wide adjustments in pricing. Affordability is considered to be 30 percent of the target household's income. The calculation takes into account the size of the household as well as its relative income; they have released data for two- and four-persons households at 80 percent and 100 percent of the area median income. Inadequate units are removed from the stock. Two affordability indexes are measured, an unadjusted one based only on housing costs for homeowners and for renter, an adjusted one that takes into account the level of amenities available in the town.

The multiple dimensions that enter into the calculation of this affordability index (housing costs, access to jobs, safety and school quality) are presented in the results. The purpose of these dimensions is to make the index useful to policy-makers as they decide where public investment should be made and what type of investment is needed depending on the place. The index sheds an analytic spotlight on the index we are reviewing by expanding the concept of access to safety and schools. The primary underlying criteria is the price/income ratio adjusted for job accessibility, quality of school and safety. Thus this measure is aimed at an affordability concept that is user-cost affordability. The strong conceptual basis to the user cost approach allows a robust comparative analysis of housing costs across locations. This method is a conceptually well-identified analysis of the affordability question in a user-costs framework that adjusts for the level of amenities.

On the question of job accessibility, the methodology developed is of interest. It points to the need to restrict investment in areas that are affordable but do not provide access to jobs. Affordable housing should be placed in areas with a high level of amenities but unaffordable. Affordable places with low job accessibility are not prime places for public investments.

4.2.4 CNN HOUSING AFFORDABILITY CALCULATOR

The CNN Affordability Calculator requires inputting information about gross income, down payment, existing debt, mortgage rate, annual property tax and homeowners insurance figures. The calculator then provides a range of house prices, loan amounts, monthly mortgage payment, taxes/ homeowners insurance and finally the total expected monthly payment that is affordable based on the household's characteristics specified above. The calculator is based on a 36 percent debt to income ratio and a housing payment to income ratio of between 28-33 percent. Estimates are based on a 30 year fixed rate mortgage. Property taxes are pre-set to \$3,500 while homeowners insurance is pre-set to \$481, but users can modify them to fit their individual situations. Private mortgage insurance is not factored into calculations.

The Affordability Calculator does not calculate an index for an area but rather, computes an estimated house price and mortgage amount a prospective homebuyer can afford based on individual characteristics. The specifications of the Affordability Calculator are such that it is not location specific, although local property tax may enter into the calculation to determine what property a household can afford.

4.2.5 WALK SCORE

The walk score is an index between 0-100 that provides a measure of the walkability of any address. While focused on walkability, the Walk Score website also provides some measure of housing and transportation affordability using cost estimates provided by Zillow.com and CNT.

The walk score is calculated by taking into account possible routes and distances to amenities, road connectivity metrics and scores for individual amenity categories weighted by importance, defined as the ability of an establishment to draw walking patrons. Amenity categories (grocery stores, restaurants, shopping, coffee, banks, schools, parks, entertainment) are ranked based on high, low and medium importance to walkability. For example grocery stores and bars are businesses that drive walkability and are therefore placed at a 3, shopping establishments and coffee shops are medium and thus a 2, and finally, "other" is weighted at 1. The walk score attempts to provide a normalized assessment of how walkable a neighborhood is. The business data used are obtained from business listings on Google, and geolocalized, road network data are obtained from open street maps, school data come from education.com and public transit data from over 200 transit agencies.

The calculation of the walk score is based on a distance decay function. "A full score is given for amenities within 0.25 miles of the origin and scores decrease with distance smoothly. At 1 mile from the origin, the amenity receives 12 percent of the full score. After 1 mile, decrease in scores fall in rate. Any amenities 1.5 miles or farther from the origin receive a score of 0." These figures are based on a walking person moving at 3mph traversing a distance of .25 miles in 5 minutes. Aspects such as intersection density and average block length can lead to a penalty of up to 10 percent of the total score.

In addition, a transit score is developed based on the proximity of transit routes and their usefulness. A score from 0 to 100 is based on the distance to the nearest stop on the route, the frequency of the route, and the type of route. "The value of a route is defined as the service level (frequency per week) multiplied by the mode weight (heavy/light rail is weighted 2X, ferry/cable car/other is 1.5X, and bus is 1X) multiplied by a distance penalty. The distance penalty calculates the distance to the nearest stop on a route and then uses the same distance decay function as the Walk Score algorithm." Transit data were obtained from transit agencies across the country.

The walk score and transit scores are accompanied by a commuter cost calculator that provides commute time and a monthly transportation cost estimate using different modes of transportation for a given route. Walk Score also provides estimates of housing costs and transportation costs for a given property, and there is a calculator to give the share of income a household would pay on transportation and housing. These are not proprietary cost estimates; instead, they use rental and ownership costs calculated by Zillow.com at the property level and transportation costs calculated at the block group level by CNT (same costs as those used in the H+T index). The calculator is set to the monthly housing cost estimated by Zillow.com (based on sale data and mortgage rates), metropolitan median household income and transportation costs for the block group in which the property falls as estimated by CNT. It is possible for a household to change these parameters to better fit its situation.

Walk Score takes into consideration a wide range of amenities into its calculation. One of the most recent updates to their methodology has also been to calculate distance based on following existing paths rather than using crow-flies distances, improving the accuracy of measurement.

4.3.6 MINNESOTA CENTER FOR TRANSPORTATION STUDIES TRANSPORTATION AFFORDABILITY FRAMEWORK

The University of Minnesota Center for Transportation Studies has developed a population specific matrix to measure the degree of affordability of transportation for different socioeconomic groups based on location within the Minneapolis-St. Paul MSA. Transportation is measured as time affordability plus financial affordability. The approach divide groups by socioeconomic status to take into account disposable income and disposable time values of individuals. Locations are divided based on categorization of the built environment and accessibility to basic daily activities. Financial affordability is defined as (needs oriented transportation expenditure)/(household disposable income) while time affordability is defined as (needs oriented travel time)/(household disposable time). "Needs oriented" is calculated based on access to work, learning, healthcare, food shopping and social cultural and sport activities. Disposable income is income after taxes, while disposable time is defined as the time left in a day after subtracting average daily time spent in work, learning, healthcare and food shopping activities. The study was done for the Minneapolis-St. Paul Metropolitan Area only and does result in the creation of an index per say. The data used have been collected between 2002 and 2010 and include the Consumer Expenditure Survey, American Time Use Data, Minneapolis Public Library, Dun and Bradstreet Business Dataset, and ACS.

Figure 6: Walk Score: Housing and Transportation Costs Calculator

Walk Score **98** Walker's Paradise
 Out of 100 633 League St Philadelphia Like 7k

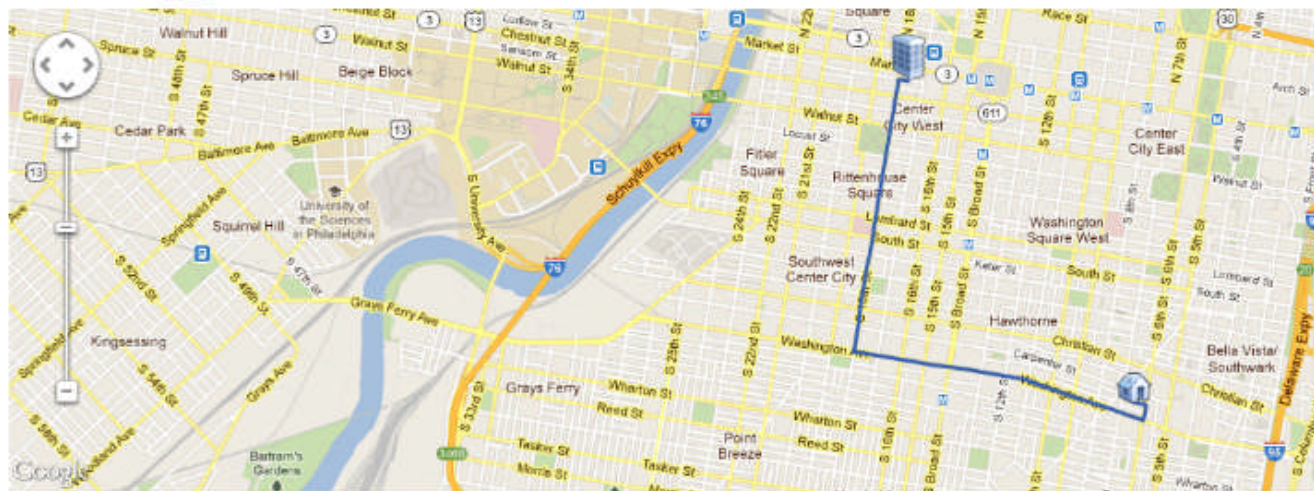
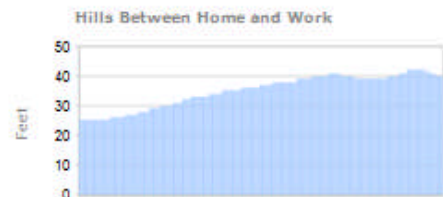
Overview More Amenities Your Commute Wharton/Hawthorne/Bella Vista

Where do you commute?

1700 market st philadelphia pa

Commuting to: 1700 Market St Philadelphia

- 39 minutes (1.9 mi)
- 16 minutes (2.16 mi)
- 9 minutes (2.16 mi)
Cost per month: \$65
- 21 minutes (1.71 mi)
Route 1 of 4 (next): 47, MFL



Walking, biking, and driving times from 633 League St Philadelphia to 1700 Market St Philadelphia. See what's nearby your [home address](#) or see what's nearby your [work address](#).

Apartment Search from Walk Score

Find nearby apartments and rentals.
 Search by Walk Score, commute time, or near transit.

Housing and Transportation Costs

Estimated costs for 633 League St Philadelphia. Type in your own numbers to personalize.

Monthly housing: \$ (Based on Zestimate®)
 Income: \$ (annual)
 Transportation: \$ (monthly)

Where did we get these numbers?

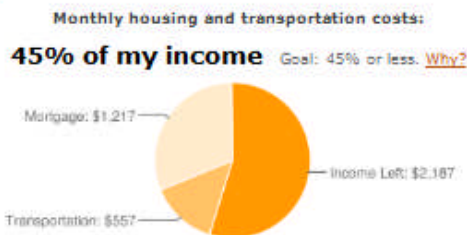


Table 10: Summary of Alternative Indexes

Affordability Index	Source	Description	Geographic Scale	Data Sources
NAR Housing Affordability Index	National Association of Realtors	Ability of median-income family to buy median-priced home based on price and effective mortgage rate. Updated quarterly and monthly and published separately for first-time homebuyers.	Nation, Regions	NAR monthly survey of existing single-family home sales. Effective Mortgage Rate from FHFB. Projected Median Family Income based on Census data.
NAHB-Wells Fargo Housing Opportunity Index	National Association of Home Builders and Wells Fargo	Percentage of homes affordable to median-income family. Updated quarterly.	Nation, MSA	CoreLogic sales data. Effective Mortgage Rate from FHFB. Median Family Income from HUD. Census data estimated property taxes and property insurance at the metropolitan level.
Affordable Housing Index	MIT's Housing Affordability Initiative	Percentage of units affordable for a household of a given size and income. Accounts for job accessibility, public schools quality, and environmental conditions.	Place (for specific New England MSAs)	Census data for rental and owner-occupied units, updated using Case-Shiller-Weiss repeat sales indexes and data on price appreciation from Zillow.com.
Walk Score Housing and Transportation Costs	WalkScore	Monthly housing costs for renters and owners (at the property level) and transportation costs (at the block group level)	Property-Block group	Housing costs based on Zillow.com's Transportation costs estimates based on CNT's estimates

4.3 H+T INDEX FROM THE PERSPECTIVE OF USERS

A survey of ten organizations that are using or have used CNT's H+T Index was completed based on contact information provided by CNT or found in studies available online that used the H+T Index or a similar concept. The survey attempted to collect the experience of those who have used the index to define possible uses of the index as well as elicit concerns and suggestions.

Among respondents some had used the index with CNT acting as consultant, producing a detailed report tailored to their needs and presenting the results of the analysis in a summary format. Others had obtained data from CNT and done their own calculations.

The respondents all worked in planning, housing or transportation agencies involved in regional master planning or deciding where to locate public investments. Respondents who had used the index with CNT acting as a consultant were in general pleased with CNT's work and found the index to be useful for visualizing housing and transportation costs in their community at a fine grain and communicating these costs to their constituency. They had not necessarily kept a working relation with CNT past this original study, but several were looking to update the study for their regions with the new data used by CNT. Those who had produced their own reports and graphic representations were also generally satisfied with the data provided by CNT, but some expressed concerns that the data sources and modeling are not always clearly laid out.

Users identified as a strength of the index its capacity to clearly visualize housing and transportation costs on maps. They also felt that the measure of cost to income ratio on which the index is based was relatively intuitive and well understood when they tried to communicate these findings to policymakers or to a general audience.

Regarding H+T's weaknesses several respondents mentioned their concern with the lack of timeliness of the data. Another common concern was the lack of transparency in the calculations done by CNT, to estimate transportation costs in particular. Several users mentioned the fact that the formulas and data are proprietary as an issue that had frustrated them when trying to understand the index.

Users were divided over whether modeled transportation costs or observed ones were preferable. Some thought that it would be preferable to use the observed data while others argued that modeled cost better captured the choice offered to residents but that the methodology should be made public. A separate point was that the coverage of the index is limited. A few respondents working for state-wide organizations also mentioned the fact that currently the index only covers metropolitan areas which prevent them from using it as a recommended tool for their grantees for example.

Finally, another concern that was raised was the use of the metropolitan median household as the representative household since for some areas, this result in a calculated ratio of Housing plus transportation costs relative to income of over 100 percent. In a related case, this concern resulted in the respondent substituting a sub-region median income to the metropolitan median income.

Most respondents emphasized that H+T filled a void in existing tools available to measure and represent housing and transportation costs at the local level. They understood that data available at this geography are sparse and would be costly to assemble if the index was not based on existing public datasets, which

would also lead to some comparability issues. Their main recommendations to improve the usability of the index for housing and transportation planning and for public investment decision purposes were to use more up-to-date information, to offer the possibility for local organizations to add their own local data to the index and make the transportation and housing portions of the index available separately. To quote one responding agency which had developed its own measure of housing and transportation costs at the neighborhood level the H+T Index served as “an aspiration for their work.” While they created their own version of the index they said that “it should be seen as a tribute to CNT’s work, not a repudiation of it.”

5.0 RESULTS

Broadly speaking, the creation of an ideal combined housing and transportation index is a daunting task. There are data availability difficulties, conceptual issues surrounding what type of household to use in the index, and numerous modeling considerations. Given these challenges, the basic framework should be improved incrementally over time. We think that the H&T Index is a good first step that currently provides useful insights along a number of dimensions. For this review, we had the opportunity to look inside the model and replicate its findings. We can confirm that the calculations and results are consistent with the description provided to the public by CNT and that the new model is at least as good as the 2008 model that had benefited from a peer review process (Haas, 2008). We have summarized the strengths and weaknesses of the H + T Affordability Index, developed a series of suggestions for ways the Index can be improved, and prepared preliminary thoughts on how the index can be used for policy purposes, including analyzing potential impact of place-based investments and the potential for incorporating use of the Index into HUD programs.

5.1 STRENGTHS AND WEAKNESSES

The strengths of the H + T Index center on the simplicity with which the results can be presented and understood and on their comparability. In addition, the index has benefitted from a rigorous approach to both data assembly and statistical modeling, although there are opportunities for significant improvement. The weaknesses include shortcomings in data transparency, availability and quality, geographic coverage, the lack of temporal element, limitations associated with a representative household, poorly developed housing cost index, as well as a number of limitations in the statistical approach. These weaknesses indicate the need to continue the development of the index rather than fundamental challenges to the overall concept or implementation. The recommendations section covers many of the weaknesses and suggests ways to overcome them.

5.1.1 STRENGTHS

Strong First Step: The H + T is a serious attempt to address the tricky questions of joint housing and transportation affordability. The index fills a void of information available to decision makers, including households, firms, communities and policy analysts.

Easy to Read Score: The Index can provide an easy to read score that can be used as a common factor that can be cross-compared between locations and properties. The unit (cost to income ratio) of the index is intuitive.

Separated Scores: The Index can be divided into two indexes. Several respondents confirmed this was an advantage, allowing them to substitute a preferred local index in some cases.

Illustrates Differences on Maps: The index makes it easy to clearly visualize housing and transportation costs on maps. The Index also illustrates differences in housing and transportation costs over space using map based illustrations.

Can guide affordable housing investment: The Index points to areas with high housing costs and low transportation costs where investment in housing could provide affordable units.

Can guide transportation investment: In certain circumstances the Index can guide investment. For example, areas with high transportation costs and low housing costs that are close to low transportation costs areas might be appropriate places for transportation investments that could provide access to affordable housing.

5.1.2 WEAKNESSES

Only available in Metropolitan/Urban Areas: The Index is not useful for statewide application, and does not measure rural access to transportation

Lacks a temporal component: Neither the housing nor transportation components of the index vary over time. Thus the indexes are necessarily dated. More importantly, tracking the evolution of costs could be very useful.

Users are unsure about the reliability of the data: Because the sources behind the Index are not clearly indicated, users have some skepticism about whether to trust the underlying data.

Users are unsure about the meaning of the Index: Because the data and formulas are not fully revealed to users, they don't have a full understanding of what is being measured by the Index and therefore are not sure how heavily they can rely on the Index.

Based on "Representative Household": As a guide for household decisions, the index works only for the "representative household," and would need to be able to be adapted to different household characteristics to inform other types of households. As a guide for transportation investments, the Index reflects equilibrium in one place at a given cost, and adding transit access would change the composition of the place, which makes general guidance on transportation investment problematic.

Different Destinations Imply Different Costs: A single index number cannot capture the variation in transportation costs that depend on the modes available and the desired destinations. Next door neighbors can have very different transportation costs. Potential households do not choose neighborhoods based on the average transportation costs for the average household, they choose based on the likely transportation costs for their household.

Based on existing households, not potential households: The housing component measures costs for existing households, which are not the costs a potential new household would face. More generally, the index is not independent of the households that live in a neighborhood.

Does not capture systematic differences between regions: The econometric model treats all regions of the country the same and does not capture systematic differences between regions.

Based on historical data for existing residents, rather than current data for potential residents: The data for the housing portion of the index are outdated and apply to existing residents rather than potential residents,

5.2 RECOMMENDATIONS

Our recommendations fall into two broad categories, technical issues and conceptual issues. Technical issues concern less central details about how the index is calculated, whereas conceptual issues focus more on what the index should be calculating. There is considerable overlap between these two areas.

5.2.1 RECOMMENDATIONS CONCERNING TECHNICAL ISSUES

Reveal the data, processes, and calculations that make up the index: One of the criticisms of people who have examined and used the H + T Index is that it lacks transparency (Abt, 2010). This lack of transparency has a cost – it makes the index less understandable, and less trusted, than if more were known about it. We recommend that CNT reveal the underlying data, processes and calculations that underlie the index. Users and interested parties who are able to see and understand that data and calculations will find the index more convincing. Further, observers might be able to offer suggestions, alternative data sources, or other improvements that will make the next iteration of the model better.

Calibrate the Index to 100: An alternative way to present affordability information is to create an index in which the values for each location are related to each other, and focus around an important threshold level. For example, H + T could be recalibrated to create an index where 100 represents the level at which housing and transportation costs are affordable to the median household.

Use housing cost data that are more up to date: The housing cost data comes from the 2009 ACS, based on 5-year averages. The current data spans a period with fluctuating housing values, which are at least three years old. Relative housing values in MSAs and between MSAs have changed significantly since that time. We recommend that the H + T index consider a more current data source, and that the data source offer regular updates.

Separated Scores: The Index should continue to be divided into two indexes. By separating the housing and transportation components of the index, users can substitute a more up to date housing affordability index for the housing portion.

Research the availability of similar data that is in the public domain: Closely related to the recommendation for fresher data is the recommendation for using data in the public domain. Using publicly available data builds confidence in the index because users understand where the underlying data come from and can access it themselves for clarification if desired. Of course, the data must be trustworthy as well.

Make it easy for local organizations to integrate their own data to the model: The H + T model uses the same method and data across all covered MSAs in the US. This commonality facilitates inter MSA comparisons and provides for consistency in data handling and calculation. However, there are instances

in which organizations are interested in a single MSA and have data for the MSA that is more precise or more current than the data that underlies the H + T index. We recommend that CNT provide an easy method for these organizations to substitute data, such as a MSA specific housing affordability index, for the corresponding native H + T data and use the rest of the H + T data and algorithm to compute the local index.

Publish regression analysis coefficients rather than just an index: Related to the idea of unbundling, users might be interested in the regression coefficients, particularly if the index is recast as a linear model. Knowing the regression coefficients would let the user calculate a custom index, for example for single people not interested in automobiles or for a family not interested in taking transit. We recommend that CNT publish the regression coefficients with enough explanation so that users can run custom predictions.

5.2.2 RECOMMENDATIONS CONCERNING CONCEPTUAL ISSUES

Calculate the index state-wide, measuring all transportation types: The H + T index calculates values for block groups in metropolitan areas, but not for housing located outside a metropolitan area. However, the concept of housing and transportation affordability is relevant in non-metropolitan communities as well. Though the number of alternative transportation modes outside metropolitan areas is smaller, the distances traveled are frequently longer, so an index of location efficiency would be valuable to people considering non-metropolitan locations. We recommend that the index be extended to cover all areas of a state.

Develop a calculator/matrix to provide information tailored to different household types: The index calculates costs for the “typical” household in a census block, which can mask large variations in the actual inhabitants of the block. The use of the typical household, as discussed, also relies on the financial choices of the residents as well as how long they have lived in the neighborhood, and does not necessarily provide information for a household considering a move to a block. We recommend that CNT build a calculator, potentially an on-line calculator that allows the calculation of an index for a variety of household types and work destination.

Consider adopting a linear regression framework for ease of understanding: The current regression framework uses a highly non-linear model. This model framework is very flexible in that it can estimate a variety of function with relatively few parameters, but it is not possible to interpret any of the individual parameter estimates. Standard economic modeling begins with a linear model, typically Ordinary Least Squares (OLS), and only moves away from OLS if there is strong reason to expect that OLS will fail in a statistically relevant way. Linear estimators like OLS are widely understood amongst economists and are used because they possess helpful statistical properties and because the coefficients can often be given real world interpretations. We understand that CNT has investigated the use of linear models, and that the linear models arrive at nearly the same index value as the currently used non-linear model with only a minimal loss in explanatory power. We recommend that CNT adopt a linear model for the index in place of the model now used. The linear model will be especially helpful for users of the index if regression coefficients are made available to users, which is another recommendation.

Use of the housing affordability concepts that are embodied in industry standards such as the NAR and HOI: The problem with the H+T measure is that it is based on historical data for existing residents,

rather than current data for potential residents. At a minimum rent and homeownership costs should be updated using regional annually available BLS data on rent, from the CPI, and price data from FHFA. In addition homeownership costs for the representative potential household, using the median regional income household, for example, should be created using standard down payment and current mortgage rates.

Create a longitudinal dimension to the index: The index could be made available for different time periods, and updated regularly in order to show the evolution in the housing and transportation affordability of a given location over time. Tracking the evolution of costs could be very useful. In addition the methodology of the index could still be modified and historical values recalculated using the most recent methodology as long as the data are preserved.

5.2.3 THE USE OF THE H + T INDEX FOR POLICY PURPOSES

The Index has several roles to play in policy setting purposes. Preliminary thoughts on policy

- The Index's usefulness depends on details and goals of policy.
- The index may be useful for helping to direct Low Income Housing Tax Credit (LIHTC) investments. Because it identifies low transportation cost areas, the Index can point to places appropriate for low income housing.
- An important limitation is that policy makers need to be very careful if it is intended to use the Index to guide transportation investments. Improved transportation typically increases housing costs, so that decreases in transportation costs are mostly or wholly offset by increases in housing costs. The Index, because it is static, cannot recognize this dynamic phenomenon. Fortunately transportation investments usually require significant, project specific review so there are other ways to address this issue.

5.3 CONCLUSION

We conducted this review to assess the strengths and weaknesses of existing measure of housing and transportation affordability indices with a particular focus on CNT's H+T Affordability Index with the goal of informing the development of HUD's HTA Index. Our report suggests areas that can and should be addressed to attain the goals of the proposed HTA index.

In this report we looked at the rigor of the methodology designed by CNT and replicated the transportation portion of the index. While there are discussions in the literature of this index, to our knowledge it is the first time that an exterior group is given the opportunity to access the detailed methodology behind the index and to replicate the findings. We find from this assessment that the results and measures correspond to these described by CNT.

From the review of the literature on indices and survey responses of users of the index and our own review of the housing and transportation side the evidence is that the H+T Index provided a measure of

affordability that was not available and can be used to meet the goals outlined by HUD as a base for its HTA Index with some changes to the methodology and data used.

REFERENCES

Abt Associates Inc. (2010) "A Methodological Review of the Center for Neighborhood Technology's 'Housing+Transportation Affordability Index.'"

Banzhaf, Spencer and Omar Farooque. (2012) "Interjurisdictional Housing Prices and Spatial Amenities: Which Measures of Housing Prices Reflect Local Public Goods?" NBER Working Paper 17809, <http://www.nber.org/papers/w17809>

Bogdon, Amy, and Ayse Can. (1997). "Indicators of local housing affordability: Comparative and spatial approaches." *Real Estate Economics*, 25(1), 43-60.

Danielsen, Karen, Robert Lang, and William Fulton. (1999). "Retracting Suburbia: Smart Growth and the Future of Housing." *Housing Policy Debate*, 10(3), 513-40.

Diaz Antonia and Maria Luengo-Prado. (Forthcoming) "The user cost, home ownership and housing prices: Theory and evidence from U.S." International Encyclopedia of Housing and Home, London, UK: Elsevier.

Fan, Yingling and Arthur Huang. (2011) "How Affordable is Transportation? A Context-Sensitive Framework." Center for Transportation Studies, University of Minnesota

Fisher, Lynn, Henry Pollakowski and Jefferey Zabel. (2009) "Amenity-Based Housing Affordability Indexes." *Real Estate Economics*, 34(4), 705-746.

Haas, Peter, Carrie Makarewicz, Albert Benedict, and Scott Bernstein. (2008). "Estimating Transportation Costs by Characteristics of Neighborhood and Household." *Journal of the Transportation Research Board*. 2077, 62-70.

Linneman, Peter and Susan Wachter. (1989). "The Impacts of Borrowing Constraints on Homeownership," *American Real Estate and Urban Economics Association Journal*, 17(4), 389-402.

Massachusetts Institute of Technology, "Measuring Housing Affordability: the housing affordability initiative housing index." <http://web.mit.edu/cre/research/hai/aff-index.html>

National Association of Home Builders, "Methodology for the housing opportunity index." <http://www.nahb.org/generic.aspx?sectionID=135&genericContentID=533>

National Association of Realtors, "Methodology for the housing affordability index." <http://www.realtor.org/research/research/hameth>

APPENDIX 1 INDEX THEORY

There are two significant issues with how CNT constructs its housing model, 1) the use of a payments approach instead of a user cost or Owners Equivalent Rent (OER) index, and 2) the lack of a regression approach to control for housing quality.

The housing portion of the H+T index is not consistent with cost of living index theory. Economists generally consider user cost and OER to be the most theoretically and practically sound measures of housing cost. The H+T measure is closest to the payments or “cash flow” approach that looks at out of pocket housing expenses. It is important to note that this method is not very popular among economists or price statisticians for many reasons. This is the method the U.S. used from the early 1950s until 1983, but is used by few countries today¹⁴ (the majority use rental equivalence). The housing cost component of the CNT H+T index shares the shortcomings of the payments approach.

One important concern is that CNT’s measure will be extremely volatile with respect to household portfolio decisions. For instance, if a homeowner sells bonds and uses the funds to pay off their mortgage their housing cost as measured by CNT would go down significantly because mortgage payments will go down. However, homeowners can go back and forth between portfolio decisions like this and affect the measured cost of living and it does not affect the actual cost of living in any way.

The CNT approach is also problematic because the index is missing a component to capture the down payment portion of the sales price of the house. Even for long-time residents, the model does not propose an adequate representation of the ongoing opportunity cost faced by existing homeowners, who face an opportunity cost of the equity tied up in a housing investment.

These issues reflect the fact that CNT’s payment-based measure of housing does not align with the commonly accepted notion that the utility a household derives from housing is due to the housing services consumed in that period, and not due to the stock of housing they own, which is essentially an investment good that produces a flow of housing services (Fisher et al., 2009). The mortgage expenses are not a measure of the cost of the flow of housing services, nor are they a measure of the cost of the stock of housing.

Even ignoring the more academic problems with the CNT measure of housing, it has problems from the standpoint of a commonsense conception of housing costs. From the standpoint of someone choosing where to live, or of policymakers trying to understand which areas are expensive to move to, the relevant housing costs should be based on the costs of a new resident. Because the mortgage costs used in the census measure are based on all existing tenants it includes the cost of mortgages that reflect house prices which may be 10 years old or more. Because house prices tend to go up over time, one result of this measurement is that neighborhoods with many long-term residents will systematically appear cheaper than neighborhoods with new residents, even when the housing cost for a new resident are identical.

There is also the issue of why CNT does not use a regression-based approach for housing as they do for transit. One reason they give is that the amount of transit a person uses is not as fixed within a block as

¹⁴ Christensen, Dupont, and Schreyer “International Comparability of the Consumer Price Index: Owner-Occupied Housing”, 2005 survey 30 nations and report only Ireland and the U.K. utilizing the payments approach.

housing is. The price and quantity of a house are also not as decomposable as for transit, and it is price more than quantity that fluctuates from neighborhood to neighborhood. Possibly there is a regression based solution to this, but predictions of housing quantity use would need to be constrained to be within the distribution of existing quantity, regardless of the level of housing use predicted by the control household.

The issue of how much quality adjusted housing stock is available in the block group illustrates a conceptual issue with any index like this. A cost of living index for relative geographic costs, like the CNT index, asks what is the minimum expense needed in area A to achieve the same amount of utility they can get in area B. This presents several issues. Like what is the base level of utility in area B? If one instead asks what is the minimum expense needed in an area to buy a particular amount of the good, then this begs the question as to what amount should be considered?

There is no right answer to this question, and it depends in part on whose cost of living is being measured. One option is to choose a particular income level, and the base level utility used is thus the maximum achievable for that income in area B. One could also choose a level of the good consumed in the base area. In this case that would be the quality adjusted amount of housing.

In the case of the CNT affordability index, the intent is not to measure the price of all block groups relative to some base case, but to measure absolute costs in each block group, which can then be used for relative comparisons. Consider the traditional COLI measure showing the ratio of expenditure functions in two areas, which shows for a given set of prices what is the minimum cost way to achieve a particular utility level U:

Expenditure function = $e(U, P)$

COLI = $e(U_x, P_A) / e(U_x, P_B)$

Where P_A is the set of prices in area A, and P_B is the set in area B. CNT instead creates an expenditure for each given area, like the following:

Indexes: $\{e(U_x, P_A), e(U_x, P_B), e(U_x, P_C) \dots e(U_x, P_N)\}$

In either case, the question of which base utility level to use must be asked. This can be a difficult conceptual and empirical issue to answer, so often the indirect utility function V is substituted, which shows the maximum level of utility achievable for a given set of income and prices.

COLI = $e(V(I_x, P_x), P_A) / e(V(I_x, P_x), P_B)$

Now instead of base level of utility, one must simply choose the base level of income and the base prices. Given some level of income and prices, the individual would choose the utility maximizing quantity of the good to consume.

When the component of the cost of living index dedicated to a single good, say housing, is under consideration, the following measurement is often used:

Index_{H,A,B} = $H_x P_A / H_x P_B$

Where H_x is the quantity of the good, housing, consumed in the base area, which in this example could be A or B. An important limitation here is that the index for this good does not capture the welfare from spending less on housing, and either saving more or spending more on other goods, that might occur in either location rather than purchasing more housing. Another way of saying this is that by assuming an individual purchases the same amount of housing in either block area, this measure does not allow for substitution into other goods. An individual buys 1,000 sf of housing in neighborhood B might only buy 750 sf of housing in neighborhood A if prices are higher in A. Assuming that they buy the full amount of housing in A that they bought in B provides an upper bound on the amount that they would have to spend in order to be just as well off in A as in B. The traditional affordability indexes get around this by using the value of the median house. Linneman and Wachter (1989), and the following on literature, estimates the optimal house, for a given household, as a function of a series of variables influencing housing demand and then compares the required down payment for that house to actual or predicted wealth for that household, and compares mortgage payment, using standard mortgage rates, to income or predicted income for that household to determine whether that household is “constrained.” That is to determine whether that household is facing a housing market that is not affordable. This creates a regression method for identifying affordability.

Another way to measure the cost of living for a particular good would be to use the utility maximizing level of housing spending in each neighborhood given a particular level of income and prices as the measure of housing quantity consumed. This would be represented as:

$$\text{Index}_{X_{H,A,B}} = X_H(I, P_A)P_A / X_H(I, P_B)P_B$$

Where X_H is the marshallian demand for housing given income I and prices P . However, this underestimates the decrease in utility from an expensive area relative to a cheaper area if the elasticity is greater than zero. If the elasticity is equal to one, then increases in price will be perfectly offset by decreases in quantity purchased, creating the appearance that the index value is 1. Or consider if the index does not present relative costs, but rather a list of absolute costs, as CNT does, like the following:

$$\text{Indexes}_H: \{X_H(I, P_A)P_A, X_H(I, P_B)P_B, X_H(I, P_C)P_C, X_H(I, P_C)P_C, \dots, X_H(I, P_N)P_N\}$$

In this case, for elasticity equal to one, it will appear as if the cost is equivalent across all areas regardless of the relative prices. For this reason, this type of index is best seen as approaching a lower bound of the relative cost of living.

This is conceptually closest to what CNT has attempted for transportation and what would be done with housing to be consistent across expenditure groups. A representative individual for each MSA is chosen, recognizing the endogeneity of individual characteristics like income to choice of city. Then the predicted level of transportation purchased can be estimated, which is an attempt at estimating individual marshallian demand for a city level representative individual.

In the context of housing, this issue presents several difficulties. For one thing, within each block group $X_H(I, P_j)$ is much more constrained choice than it is for housing because housing available is a highly fixed stock. This means that within each block group, an individual must choose $X_H(I, P_j)$ subject to the constraint that X_H is a level of housing that is actually available in that neighborhood.

In addition, the amount of housing purchased within a block is to a large degree a function of block specific prices, which can be seen in the fact that X_H is a function of I and P_J , the block specific price of housing. But that high prices would lead someone to consume a small amount of housing, and thus perhaps spend very little on housing overall does not fully reflect the loss of welfare due to higher housing prices.

However, as discussed above, using a fixed amount of housing as the baseline in each neighborhood as an upper bound on the cost of housing, that ignores the fact that one can minimize welfare losses by spending less on housing in less expensive neighborhoods. This issue is a fundamental one in the discussion of the meaningfulness of creating neighborhood-level index of housing affordability. Neighborhoods are areas within MSAs with similar quality housing stock and amenities, MSAs are the housing market, neighborhoods are not. Rather they are collections of similar housing. Choosing among neighborhoods within an MSA and saying one is more expensive than another is simply saying that the housing (and amenity level) there is better located, bigger and better.

APPENDIX 2 SURVEY RESPONSES¹⁵

User 1

User 1 told us that CNT H+ T index was and still is ground breaking. Accessibility is an enormous asset; it helps bridge knowledge gaps between consumers and industry, and does an excellent job of using the available data. In a perfect world, you could have more detailed and exact calculations. The user noted that CNT separately will do customized studies for local communities that are more detailed. CNT does go deeper in these communities, and they are moving in this direction. The user felt that the CNT H+T Affordability Index is primarily limited by the data available, as opposed to the execution of the model. It fulfills its intended use, while relying on available data. One possible area of improvement is to add other expenditures to the model, such as food and energy costs.

User 2

User 2 described CNT H+ T Affordability Index as a fairly robust tool that communicates complex ideas simply. Primary concerns are related to oversimplification, and inability to satisfactorily account for income, particularly in poorer neighborhoods where the index shows H and T costs at 80% of income or more. Similarly, the user felt that there was a false sense of prediction and it is important to understand that these are expected H+T costs, and actual costs, particularly in relation to income, will vary greatly on a household to household basis. User suggested that this limitation might be overcome using household expenditure data, as opposed to the self-reported income data from the census. User strongly encouraged the idea of separating the housing and transportation data, commenting that it is one of the most frequently requested pieces of information when presenting the results of the index to clients.

User 3

User 3's organization received H+T index data from CNT. They received the housing costs and transportation costs as separate numbers. The user has been using it for almost two years. They do not conduct statistically robust analysis with them but map the data obtained to visualize housing and transportation costs around several planned transportation corridors in a metropolitan area. They use the data to visualize the housing and transportation affordability level in these corridors now and how they compare to the rest of the metropolitan area. The user felt that the value calculated by CNT fits the reality with core area of the metropolitan area well served by high-frequency transit having low transportation costs and exurbs showing low housing costs but high transportation costs. They also overlaid their own data on subsidized housing on top of maps showing CNT's data.

The value of the index is similar to walk scores or carbon footprint calculators: the user plans to use it essentially to illustrate local housing and transportation costs in an effort to create community engagement. The user thinks it can be used to show the value of the new transit lines being planned in areas that currently show high transportation costs. Data has not been used for this purpose yet because, as of now, they only have access to 2000 data, and they are cautious to publicly display outdated data. The user is waiting for the updated index-based ACS data but thinks that while they will be more recent they will still rely on data collected in 2005 which is already 7 years old; there has also been doubts expressed about the use of the ACS data at the block group level, arguing that calculating the index at the census tract level

¹⁵ Survey respondents were told that we would not attribute their quotes but that we would list their names as having been contacted for this review.

might still be fine-grained enough for neighborhood planning while being more robust in terms of the sample size.

Originally, some of their funders wanted them to look at indicators other than housing costs to guide their action, which is how they became interested in using CNT's H+T Index. Currently they received a sustainable communities grant and have included the index as one of the metrics they will use to evaluate change in a 1 mile buffer of the transportation corridors they are working on. Additionally, the organization claims to use CNT's data to develop an online housing and transportation calculator similar to what CNT developed. It is important for the user that this calculator be able to have individual household characteristics inputs. One of the things the user would like to do is to integrate their own data on rent (based on their own portfolio of rental units) and MLS listing to the model in order to have more up-to-date housing cost data.

User 4

User 4 has used the transportation part of the index for master county planning, transportation/greenhouse gas, and submitting grants/plans to NPO. For the geographic scale that the index covers, it provides another depth at the block group and has a simple interface with good usability. Additional features are financially and hence pragmatically unfeasible. A shortcoming is its lack of data for more rural areas which has made it difficult for him to make comparisons between rural and urban townships in the user's county of interest.

User 5

User 5 had mostly favorable comments to make about the CNT H+ T housing affordability index; it was innovative when the user began using it in 2008, and fundamentally changed the way that the organization looked at planning for the region of interest. The user's major concern was that the data sources and modeling are not always clearly laid out. The user trusts that they are accurate, but when using the Index and people question the validity of the data, it becomes harder to get points across. For the user, more indexes like this one can always use more accurate, but the user understands that there is a limit, and would not want to sacrifice the performance of the overall model to tighten a few aspects of the data here and there. Overall the user has been very happy with the performance of the model.

User 6

User 6's organization experience with CNT's H+T index has been mostly favorable. They received funding to hire CNT as consultant to conduct a study. The user said the findings of this study were important, in particular to show how high transportation costs actually are in certain areas of the region of interest. The user believes the study did not lead to implementation plans by individual municipalities and that the lack of funding made it unlikely that they would update the study but that the findings have been quoted by other government entities. Information about transportation costs at a local level are particularly important to them as a State law requires that they integration of transportation and housing plans. The issue with the data provided to them by CNT was that they were outdated.

User 7

The H+T Index is currently a tool that is recommended to Affordable Housing Developers applying for Low-Income Housing Tax Credits (LIHTCs) through their organization. The Qualified Allocation Plan which governs the allocation of LIHTCs quantifies access to services, jobs and transportation in other ways without using the H+T Index (Live Near Work incentives, for example). The H+T Index is not used or

endorsed in an official capacity. It is simply recommended to developers as one way for them to measure the accessibility of their proposed developments to transportation. For the last two years, H+T has been recommended in the Qualified Allocation Plan.

User 7 felt that by not controlling for differences in housing quality, the H+T Index appears to discount personal choice altogether, ignoring the choices that consumers make all the time in their housing choices (i.e. higher quality housing, neighborhood preferences, lifestyle preferences, access to family, etc.)

Another issue with the index from this user's perspective is that in order to inform prospective residents, the housing cost data should be more current, and regularly updated than the census data. Perhaps ACS data or AHS data can be accessed. Additionally, current housing costs (which are constantly changing) should be used for this to be a useful tool for public consumption.

This user also would prefer to have a measure based on observed transportation choices (for example, driving part way to work and parking near public transportation, or use of taxis, or carpooling) rather than modeling them. User 7 also noted that the H+T Index, at least in the state in which the organization is located, does not take advantage of the existence of comprehensive non-metro databases/inventory of local transportation systems/options available through grant and university-funded research and technical assistance.

User 8

User 8's organization developed its own measure of housing and transportation costs based on CNT's publically available documentation. They used their travel demand model to determine transportation costs, and analyzed affordability both in terms of the median income household for the region (following CNT) and in terms of affordability to current residents. They found out that virtually all of the city is affordable to the region's median income household, but much is still not affordable to people who currently live in the city. In addition they simulated effects of various gas price levels on affordability.

They used their analysis primarily to inform discussion about commute costs and development decisions. And intend to use this sort of analysis for public outreach as they create a Regional Plan for Sustainable Development with funding from a HUD Sustainable Communities grant.

User 8 mentioned that it found that CNT's estimates of VMT from a model which treats VMT as a function of residential density, block size, job density, and some other factors seems perfectly reasonable and appropriate. But since they had estimates specifically tailored to their region, they decided to use them. It pointed out that may be that the relationship between residential density and VMT is different in their city than it is in the rest of the country, or even within their region, the relationship might be different between other cities. Since they had tools available to generate region-specific numbers, they used them.

In addition, they used LEHD for commuter flows because this is observed data rather than modeled data. User 8 mentioned that in his view, observational data is always better than modeled data, if it is available, which is why they used LEHD data for commute trips. But they did not limit their estimates of VMT to commuter trips and for other types of trips they used their travel demand model. User 8 did not think that focusing on job access is a weakness of the CNT model, just a different way of looking at the issue.

Overall, user 8 concluded that CNT's work is very good and very important. It has served as an inspiration for their work. For him, the minor modifications they made should be seen as a tribute to CNT's work, not a repudiation of it.

User 9

User 9 used the H+T Index as part of a long term envisioning effort in a fast growing region. The H+T Index was a tool to make people aware of the different costs that could be associated with the development occurring in the region. In their vision they had to make policy recommendation for transportation and land use strategies for currently urban, suburban and rural areas. The idea behind the index is very important and was effective to communicate to people the effect of various growth patterns. They would show them maps for the region at large and refers them to the online platform to have a more detailed look for their particular area. User 9 did not have access to a detailed report for the region or to CNT's expertise. They only used the information available online in working with housing advocates and transportation experts to develop the idea that it is important to be tracking affordability in including housing and transportation. Generally they found that the index reflects the reality of their region, except in certain fringe areas. But used it a regional scale and at this scale, it conveys the broad picture efficiently.

Good tool but wish it was made available in reports tailored to local areas that incorporate community data. They found it to be a simple index to use that made it easy to communicate and convey message across with decision makers and community that housing and transportation costs vary tremendously from one area to another. Combination of the 2 items recognizes the real costs of location in a way that was not available before. One of the weaknesses of the Index for their use was in areas with a lot of growth and development happening where using census data from 5 or 10 years ago is not as relevant a measure of housing costs. They also would find it valuable to have options to customize since households in particular areas within the region might be very different in terms of characteristics (in particular size, age). For long term planning purposes, households being changing, it would be good to be able to adapt the model to take these changes into considerations. In certain areas where household are heterogeneous (many single households and large households within the same neighborhood) a model calibrated on the median does not tell full story.

If possible user 9 recommended making a basic level of the Index more readily available online for a few household types as it would help a lot of policy discussions. In addition it would be useful for them to have more granular and customized analysis broken down by more detailed household types as it would be useful to create different scenarios.

User 10

User 10's organization is using the H+T index to evaluate the impact of transit and land use scenarios on neighborhood affordability and to evaluate growth over the next 40 years. They have been using it since CNT published their first report of 28 metro areas in 2006 in complement with Walk Score and more traditional analysis.

According to user 10, the Index relies too heavily on residential density which by itself should have little impact on transportation costs. User 10 did not think that the index needs to be published in separate pieces. They found the index to be fairly reliable and most inconsistencies arose from coding errors of the land use data CNT got from local jurisdictions in their region as they worked directly with them on a customized report.

User 10 recommended eliminating residential density and using a better proxy for such as WalkScore or similar index should be tested. In addition the index uses "as the crow flies" geometry as opposed to a network analysis of transportation options which is a weakness. In addition for this user, the measure of transit access could certainly be improved. They don't take into account geographic barriers, and they

don't include whether the transit actually gets you to where you need to go. An interesting area of study would be to compare the Bookings' study on transit access to jobs.

The main criticism concerned the use of census data on ownership costs which according to user 10 relies on data from owners who may have bought their houses several years earlier when interest rates and costs were different. As an example a neighborhood with a new metro station in DC saw the sharpest increase in prices through 2007 and the mildest decline in prices since. Census data on ownership would miss that, but census data on rental would capture some of it. Unfortunately to get census data down to the tract level you need a five year sample of the ACS with only CPI trend multipliers to bring it current. Sales data would reflect the current cost much more accurately.