Walkability and Commuting Behavior in Maryland

Lillian Byrne

May 10, 2022

Abstract

This paper studies the effect of differences in walkability, the characteristics of the built environment that influence the likelihood of walking as a mode of travel and commuter mode choice. This research is relevant to urban planning, environmental policy, and efforts to curb urban sprawl and automobile use. The existing literature finds a small but significant impact of walkability on commuter mode and mode choice. I examine the impact of walkability on commuter mode choice (e.g., driving or walking to work) and commute times in Maryland using cross-sectional data from 2019. I quantify the built environment using a walkability index, which has not been widely done in the existing literature. I find that more walkable areas in Maryland are associated with decreased car ownership and use.

I would like to thank Professor Maureen Cropper for her guidance as well as Professor Nuno Limão, Jaehong Choi, David Byrne, and the 2022 University of Maryland Economics Honors Thesis Cohort for their support throughout this year. I would also like to thank the EPA's Office of Smart Growth and the creators of the EPA's Walkability Index.

I. Introduction

The relationship between the built environment and travel behavior is relevant to issues associated with urban sprawl and automobile use. Urban sprawl is the tendency toward lower city densities as city footprints expand (Nechyba and Walsh, 2004). While rapid urban growth is an indicator of a robust economy and a growing population, it is also associated with various market failures, including the social cost of road congestion, pollution, loss of open green space, and unrecovered infrastructure costs from less compact development (Brueckner, 2000; Nechyba and Walsh, 2004). Moreover, transportation is the largest end-use sector for carbon emissions in the United States and the only sector where emissions have consistently increased in the past decade (Energy Information Administration, 2019). Policies that reduce personal automobile use and ownership and encourage low-carbon transportation could be an important solution to the climate crisis (Environmental Protection Agency, 2021).

I use walkability to measure urban form. Walkability, or how "walkable" a neighborhood is, depends upon characteristics of the built environment that influence the likelihood of walking used as a mode of travel (EPA, 2021). This analysis quantifies the built environment using the Environmental Protection Agency's (EPA) walkability index. The index reflects a weighted average of indicators including distance to the nearest transit stop, intersection density, and housing and employment land-use mix. I examine the impact of the built environment on commuter mode choice (the choice to drive, walk, or take public transit) and car ownership in Maryland using cross-sectional data from 2019. I hypothesize that more walkable areas in Maryland, as defined by the walkability index, have a smaller fraction of people who drive to work and less car ownership.

Using data on over 1,300 census tracts in Maryland, I estimate that impact of walkability, measured at the census tract level, on the odds that commuters drive (or take public transportation) to work and the odds that a household owns no (or three or more) cars. To control for the possibility that people self-select into more walkable census tracts, I instrument for the walkability of the tract using measures of the age of the housing stock in the tract. My results show that increased walkability is associated with a decrease in the likelihood that someone owns two cars or three or more cars and a significant increase in the likelihood that a household owns no cars. The fraction of people driving to work decreases by 0.17 with one standard deviation increase in the walkability index and the fraction of individuals walking to work increases by 0.03 and the fraction of people taking public transit increases by 0.05.

II. Literature Review

Existing literature finds a small causal impact of the built environment on commuting and travel behavior. Cervero and Kockelman (1997) developed measures of the built environment using what is known in urban planning as "the 3Ds" (density, diversity, and design). The authors looked at 50 neighborhoods in the San Francisco Bay Area, considering thirteen built environment variables as measures of density, diversity, or design. They conducted a basic factor analysis and found that density, land-use diversity, and pedestrian-oriented design discourage travel by automobile in a "fairly marginal" albeit statistically significant way.

Bento, Cropper, Mobarak, and Vinha (2005) show that measures of urban sprawl and transit availability have modest effects on commuter mode choice and vehicle miles traveled (VMT) by looking at the effects of city shape and public transit supply in twenty-six cities in the United States. The authors use data from the 1990 Nationwide Personal Transportation Survey to create city-wide measures of sprawl and transit availability. They use this data to estimate a set of models explaining whether a worker drives to work, and the number of vehicles owned, and miles driven per vehicle for households. My analysis differs because it does not use VMTs as a dependent variable and I use a walkability index to measure the built environment, rather than city shape and public transit supply.

One issue in analyzing the built environment is the number of potential measures of urban form and the challenge of capturing diversity in urban form without introducing collinearity. In the papers discussed here, measures of the urban form include the shape of the road network, city shape, road density, patterns of residential land use, population centrality, the distribution of employment, and balance of commercial property versus housing. Walkability as a measure of urban form originated in the public health literature. Frank et al. (2005) first introduced the concept of a walkability index to study the relationship between physical activity and the urban environment, weighting variables for land use, residential density, and intersection density into one walkability index. Most "walkability indices" include several indicators of walkability weighted differently. Frank et al. (2010), for example, included net residential density, retail floor-area ratio, and land use. The Environmental Protection Agency (EPA) maintains a walkability index created with four measures of walkability: employment and household entropy, employment entropy alone, street intersection density, and distance to the nearest transit stop. Entropy measures employment and housing diversity. Ares with multiple kinds of employment (retail, commercial, service, etc.) and multiple kinds of housing (single-family homes, apartments, condos, etc.) have higher entropy scores out of 1. The EPA's land-use entropy is a measure of land use mix which considers the relative percentage of housing and employment land use within a census tract. Higher levels of entropy mean a higher mix and values range from

0 to 1. To my knowledge, the EPA's walkability index has not been used in the economics literature to assess travel behavior.

Measures of commuter behavior vary depending on data available for a given project and the level of analysis. However, the most widely used method of measuring travel behavior in the literature is VMTs. That said, other methods include car ownership (Chen et al. 2008), travel mode choice (Scheiner and Holz-Rau 2007), and weekly steps walked at the individual level (Wells and Yang 2008).

Various specification challenges arise when determining a causal relationship between the built environment and driving behavior. People more averse to driving may self-select into neighborhoods or areas that are more walkable or transit-oriented. People's preferences for living in more walkable areas mean they are less inclined to own several cars or drive to work. This creates an endogeneity problem that may confound a causal conclusion. Cao et al. (2009) surveyed various methods to control for endogeneity and recommend using a longitudinal model with control groups if data allow. While this method may be ideal, it requires many years of longitudinal data, as changes in the built environment happen slowly over time (Bento et. al, 2005). I do not have access to an appropriate longitudinal dataset for Maryland; instead, I use cross-sectional data and instrument for the walkability index.

To correct for endogeneity, Kitamura et al., (1997) created statistical control variables to examine the effects of land use and attitudinal characteristics on travel behavior for five San Francisco Bay Area neighborhoods. The authors regress socio-economic and neighborhood characteristics against the number and proportion of trips using various modes of transit. They find that their measures of travel behavior show neighborhood characteristics add significant explanatory power to their model once they control for socio-economic neighborhood differences. The authors look at variations in residential density, public transit access, land use, and sidewalk availability and find that these measures are significantly associated with transit mode choice. Additionally, the authors construct several attitudinal factors from survey results: pro-environment, pro-transit, suburbanite, automotive mobility, time pressure, urban villager, and workaholic. They find that the attitudinal variables explain a high proportion of the variation in the data. This suggests the need to control for self-selection in my analysis, as attitudes influence an individual’s choice of where to live and their commuting behavior.

Boarnet and Sarmiento (1998) use travel diary data for a sample of 769 southern California residents to look at the relationship between land-use patterns and neighborhood level and non-work travel. They model the number of non-work automobile trips a person makes using both socio-demographic variables and land-use characteristics. To control for endogeneity and self-selection within their sample, the authors chose three variables as instruments for the built environment: share of the population that is African American, share of the population that is Hispanic, and share of housing built before 1960. The authors put forward these instruments because they theorize that they are likely to be correlated with land-use patterns, but they are non-transit neighborhood amenities and exogenous to transportation choices. Similarly, I use instrumental variables, although I avoid using the demographic variables Black and Hispanic as instruments as they are likely to be correlated with car ownership and use in Maryland.

To control for endogeneity, Handy et al., (2005) use a longitudinal design to control for attitudes that stay constant over time and other constant variables, following a sample of 1,682 individuals and looking at VMTs per week and changes in driving behavior as neighborhood characteristics. Their cross-sectional analysis finds attitudinal influence on driving distance and their longitudinal analysis shows that the built environment can influence VMTs. My analysis differs because I do not use VMTs as a dependent variable or use longitudinal data. This paper is important because it found that neighborhood characteristics impact driving behavior.

III. Data

The data used in this analysis are from the American Community Survey (ACS) 2019 5-Year estimates and the Environmental Protection Agency's (EPA) Smart Location Database (SLD), which uses cross-sectional data from 2019. The dependent variables (mode choice and car ownership) are from the ACS. Characteristics of the built environment and the walkability index are from the EPA Smart Location Database, a publicly available database created by the EPA Office of Smart Growth. This dataset includes the walkability index measured at the census block group level as well as information about the distance to the nearest transit stop, intersection density, and land-use mix (i.e., residential, or commercial), all measured at the census block group level. I aggregate data from the SLD from the census block group level to the census tract level to make it compatible for analysis with the data from the ACS which is measured at the census tract level. The data represents the average of individuals in a census tract.

The walkability index includes the average distance to the nearest transit stop and intersection density. The EPA calculates a walkability score for each of the component measures by putting block groups into four quartiles for the weighted sum of each variable. The block groups are ranked from one (lowest relative walkability) to 20 (highest relative walkability) based on their values within the quantiles. The ranked scores are then weighted using the below formula.

$$\text{National Walkability Index score} = \frac{w}{3} + \frac{x}{3} + \frac{y}{6} + \frac{z}{6}$$

w = block group's ranked score for intersection density

x = block group's ranked score for proximity to transit stops

y = block group's ranked score for employment entropy

z = block group's ranked score for employment and household entropy

Image 1 shows the distribution of National Walkability Index Scores for urban census tracts in Maryland and indicates that the data is bi-modal, clustered around 6 and 14, although most urban census tracts have a walkability score of at least ten or more out of twenty. I dropped 101 rural census tracts and chose to use all urban census tracts to narrow the scope of the analysis and control for the differences between the very rural areas of Maryland and urban and suburban areas. Here the urban census tracts are all non-rural census tracts, so suburban census tracts are included.

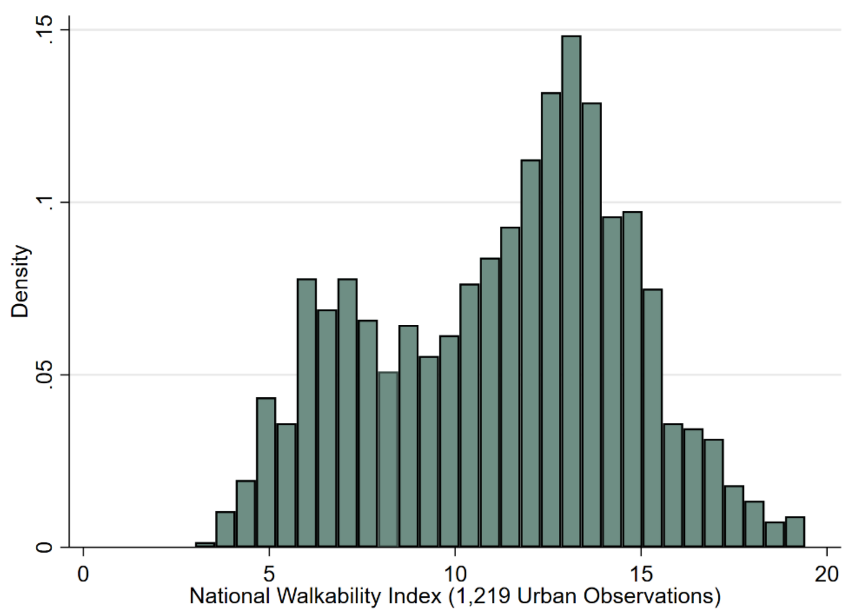

Image 1. Distribution of walkability scores for all urban census tracts in Maryland.

In describing their walkability index, the EPA cites research by urban planners Reid Ewing and Robert Cervero finding that the impact from intersection density, land use mix, and proximity to transit were all significant in explaining how frequently people walk and similar in magnitude. According to the EPA's walkability index user guide, to account for the impact of the built environment on walkability, the variables are weighted as follows: one-third to each of the three categories of street intersection density, land use mix, and proximity to transit. The land use mix category is divided into two categories to account for the two different techniques of measurement: employment entropy and employment and household entropy, each weighted by one-sixth.

Table 6 in the appendix gives a detailed description of each variable used in this analysis. Three important variables are the components of the walkability index: distance to transit, intersection density, and land use. Distance to transit represents the minimum walking distance in meters between the 2010 population-weighted census block group centroid and the nearest transit stop of any route type. EPA used a custom geoprocessing model script that selects census block group centroids and identifies all transit stops within a 1,200-meter straight-line radius. Intersection density is calculated by creating a weighted sum of component intersection density metrics. Auto-oriented intersections, intersections without crosswalks or sidewalks, have zero weight to reflect that, in many instances, auto-oriented intersections are a barrier to pedestrian and bicycle mobility. Employment entropy measures the diversity of eight different employment types (office, retail, industrial, service, entertainment, education, health, and public sector) in a block group on a 0 to 1 scale. Like employment entropy, housing and employment entropy measures diversity of employment and includes housing type.

Although the walkability index captures several features of the built environment and gives us an idea of urban form in each census tract, the process of constructing the index introduces several issues. For example, one of the three variables in the walkability index may be biasing the results in a certain direction or misrepresenting the effects of the walkability index. If one variable of the three in the index (intersection density, distance to nearest transit stop, and land-use mix) has more explanatory power than the other three, the results based on the index alone may be less dependable as the index weights them equally. To see how the individual components of the walkability index impact mode choice and commuter behavior, I perform the instrumental variable analysis for each component of the walkability index.

Table 1 shows summary statistics for urban census tracts in Maryland. The data represents the average of individuals in a census tract. Table 3 in the appendix shows summary statistics for both urban and rural areas. The mean walkability score is 11.29 out of twenty while the mean fraction of people who drive to work is 0.84. Most individuals own one or two cars while the share of people who own no cars is 0.11 and the share of people who own three or more cars is 0.19.

Finally, the dependent variable data from the ACS is the binned fraction of people who drive to work, own one car, take public transit, and so forth. To address this problem, I generate log-odds variables using the fraction of people in each census tract that engages in a particular behavior. To construct the log-odds variable, the fraction of persons driving to work (p), for example, was divided by the fraction not driving to work (1-p) and the logarithm of the odds of driving to work, $\log \frac{p}{1-p}$, calculated. 108 tracts where p was equal to zero or one were dropped.

Table 1. Urban Census Tract Summary Statistics					
Variables	Obs.	Mean	SD	Min	Max
Walkability Index (0 to 20)	1,219	11.29	3.45	3.01	19.42
Meters to transit	1,219	702.41	385.31	0.00	1,207
Intersections within ¾ mile radius of population-weighted centroid.	1,219	97.69	80.78	3.71	696.7
Land use (employment)	1,219	0.52	0.17	0.00	0.89
Land use (HH and emp.)	1,219	0.49	0.17	0.02	0.92
Median Income	1,219	88,874	42,398	11,736	250,000
Jobs within 45-minute drive	1,219	126,353	70,586	2,220	294,352
Black	1,219	0.35	0.32	0.15	0.98
Hispanic	1,219	0.13	0.13	0.01	0.96
Population sixty-five plus	1,219	0.05	0.21	0.00	0.16
Education (bachelors or more)	1,219	0.39	0.21	0.04	0.94
Fraction before 1960	1,219	0.01	0.02	0.00	0.09
Drove	1,219	0.84	0.13	0.36	1.00
Drove Alone	1,219	0.71	0.13	0.43	0.94
Carpooled	1,219	0.09	0.05	0.00	0.46
Public Transit	1,219	0.13	0.11	0.01	0.62
Walked	1,219	0.03	0.04	0.00	0.47
Biked	1,219	0.01	0.02	0.00	0.06
Other	1,219	0.03	0.02	0.00	0.12
Worked from home	1,219	0.05	0.04	0.36	0.29
No cars	1,219	0.11	0.13	0.00	0.83
One car	1,219	0.34	0.12	0.01	0.81
Two cars	1,219	0.36	0.12	0.09	0.67
Three plus cars	1,219	0.19	0.12	0.16	0.69
Commute time (minutes)	1,219	16.34	7.11	0.05	61.20

N is the number of urban census tracts in MD, data represents the average of individuals in a census tract. Zeros represent no commute time or zero people worked from home, etc. They are dropped for the log-odds transformation. Data trimmed for outliers above the 99th percentile and below the first percentile.

IV. Methodology

I estimate models to explain travel behavior using Ordinary Least Squared (OLS) and Two-Stage Least Squares (2SLS), using an instrument for the walkability index. The instrument is necessary to deal with endogeneity in the model. The estimation equations below show the hypothesized relationship between walkability and the likelihood that someone drives to work (Y = likelihood that an individual drives to work). I hypothesized that the likelihood of driving is increasing in income and education, decreasing in walkability, the fraction of people who identify as Black and Hispanic, and job density.

Ordinary Least Squared Equation:

$$Y_T = \alpha - \beta_1 W_T + \beta_2 \delta_T + \beta_3 \varepsilon_T - \beta_4 \eta_T - \beta_5 \theta_T - \beta_6 \pi_T + \mu$$

Two-Stage Least Squares Equations:

$$First\ Stage: W_{Tract} = \alpha - \beta_1 \gamma_T + \beta_2 \delta_T + \beta_3 \varepsilon_T - \beta_4 \eta_T - \beta_5 \theta_T - \beta_6 \pi_T + \mu$$

$$Second\ Stage: Y_{Tract} = \alpha - \beta_1 \widehat{W}_T + \beta_2 \delta_T + \beta_3 \varepsilon_T - \beta_4 \eta_T - \beta_5 \theta_T - \beta_6 \pi_T + \mu$$

Dependent variables = Y

Walkability Index = W

Fraction of Homes Built Before 1960 = γ

Education = δ

Median Income = ε

Fraction Black = η

Fraction Hispanic = θ

Job Density= π

Here, the explanatory variable, the walkability index, may be correlated with an unobserved factor in the error term. People may choose their residential location because of their commuting preferences. This means land-use variables could be correlated with the error term. For example, if people self-select into a particular neighborhood because it is more walkable, and they may prefer walking over other means of transportation. I test for endogeneity using the Hausman specification test for each instrumental variable regression and testing the null hypothesis that the walkability index is exogenous. The results of these tests are shown at the bottom of Table 3. The results show that I can reject this null hypothesis at least the 95% confidence level for every equation and I can conclude that the walkability index is endogenous.

To address endogeneity, I use an instrumental variable representing the fraction of homes built before 1960. I selected this variable because it has been successfully used in similar analyses and correlated with current land-use patterns in Maryland, and therefore correlated with walkability. The fraction of homes built in Maryland before the 1960s is also a non-transit related neighborhood amenity and therefore plausibly satisfies the exclusion restriction. I incorporated the fraction of individuals older than sixty-five as a secondary instrument to test for overidentification and found that both instruments are exogenous.[1] The Basmann statistic tests the null hypothesis that the instruments are uncorrelated with the error term. If you cannot reject the null, the instruments are exogenous. The p-values shown at the bottom of Table 3 are all too large to reject the null hypothesis, meaning that the instruments are exogenous. There is one exception for the regression estimating the effect of walkability on the likelihood that someone walks to work where the p-value is 0.04. However the instruments are appropriate for the other outcomes.

[1] Note that age is a demographic variable and likely to be correlated with commuting behavior to some extent. Before incorporating the fraction of people 65 or older as a secondary instrument, I ran regressions with age as a covariate to test the relationship between my dependent variables and age. I found that age was largely not significant in these regressions.

V. Results

The results of both the OLS analysis and the IV analysis show that more walkable census tracts are associated with an increase in the likelihood that someone walks, takes public transit to work, and owns no cars or one car. The likelihood that someone drives to work, owns two cars or three or more cars is inversely related to walkability. All these results, excluding the impact on owning two cars, are statistically significant at the .01 level. These results give us a good idea of the basic relationship between walkability and commuting behavior. Without controlling for endogeneity, there is some relationship between commuting behavior and walkability and walkable areas are associated with less car ownership and use.

The OLS results also indicate the importance of the control variables. Education and median income are significantly related to all measures of commuting behavior. Education is correlated with walkability and the direction of the relationship between education and walkability is the same for all dependent variables. On the other hand, income is negatively correlated with walkability and in census tracts where median income is higher is appears that people are more likely to drive to work and own at least two cars. Education appears to be correlated with commute mode choice and car ownership and the direction of the relationship suggests that more educated people are less likely to drive or own two or more cars. On the other hand, income appears to be negatively correlated with walking and public transport use: in census tracts where median income is higher is appears that people are more likely to drive to work and own at least two cars. This result may be because of wealthy suburban areas in Maryland that are not very walkable but have high median income. Suburban census tracts in of Chevy Chase, Potomac, and South Kensington, all have a median income level of at least $175,000 while the median income for all of Maryland is $88,874. Both variables suggest that median income and education level help explain some of the variation in car ownership and use across Maryland. It seems that education level decreases the likelihood that someone drives or owns a car while income increases the likelihood that someone drives or owns a car.

Table 2. OLS Results

Variables	Drove	Walked	Public Transit	No Car	One Car	Two Cars	Three+ Cars
Walkability Index	-0.04***	0.09***	0.06**	0.09***	0.05***	-0.01	-0.05***
	(0.01)	(0.04)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Education	-1.92***	2.14***	1.61***	1.12***	1.01***	-0.35***	-2.82***
	(0.14)	(0.25)	(0.20)	(0.23)	(0.11)	(0.09)	(0.15)
Log of median Income	0.62***	-1.36***	-0.42***	-2.29***	-0.81***	0.89***	2.18***
	(0.06)	(0.11)	(0.089)	(0.10)	(0.049)	(0.04)	(0.07)
Black	-1.02***	-0.059	2.29***	0.53***	0.29***	-0.40***	-0.49***
	(0.07)	(0.14)	(0.11)	(0.12)	(0.06)	(0.05)	(0.08)
Hispanic	-1.01***	-0.12	2.59***	0.26	0.42***	0.13	0.51***
	(0.149)	(0.267)	(0.218)	(0.254)	(0.123)	(0.0992)	(0.170)
Job Density	-0.05***	-0.14***	0.28***	0.031	0.01	-0.04***	-0.15***
	(0.02)	(0.04)	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)
Constant	2.68***	-1.33***	-7.39***	-0.15	-0.22	-1.74***	-2.66***
	(0.20)	(0.36)	(0.323)	(0.34)	(0.162	(0.13)	(0.22)
Observations	1,207	956	1,142	1,209	1,209	1,209	1,209
R-squared	0.508	0.309	0.634	0.624	0.384	0.626	0.652

*Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

There is a significant relationship between demographic variables and the dependent variables. Black is significantly related to the likelihood that someone drives to work, takes public transit, owns no cars, one car, two cars, or three or more cars. Hispanic is significantly related to the likelihood that someone drives to work, walks to work, takes public transit, owns no cars, two cars, or three or more cars. These variables suggest that the demographic makeup of a census tract helps explain some of the variation in car ownership and use across Maryland. It appears that in census tracts with larger Hispanic populations, individuals are less likely to drive to work but more likely to own cars while in census tracts with larger Black populations, individuals are less likely to drive to work and less likely to own cars.

Finally, job density, measured by the number of jobs within a 45-minute commute of the population-weighted centroid of a given census tract, seems to decrease the likelihood that someone drives to work and owns multiple cars. This variable captures the density of jobs in each census tract by taking the log of the number of jobs within a 45-minute drive of the population-weighted centroid in each census tract. It appears to be negatively related to the likelihood that someone drives and walks to work and the likelihood that an individual owns one car, two cars, or three or more cars. There appears to be a positive relationship between the likelihood that someone takes public transit or owns no cars and job density. While job density and the walkability index are likely correlated, the job density variable measures how dense jobs are in each census tract while the land use components of the walkability index measure the diversity of different employment and housing types.

The IV results show similar relationships, although the magnitudes of the coefficients on the walkability index increase. The Hausman test of the null hypothesis that walkability index is exogenous finds that for all IV regressions the walkability index is endogenous at the 0.05 level. Additionally, the Basmann test testing the null hypothesis that the instruments are uncorrelated with the error term finds that for all IV regressions at least one instrument is valid. These results are shown at the bottom of Table 3. The IV results also indicate that there is a negative and significant relationship between the likelihood that someone drives to work and owns one car, two cars, or three cars and walkability. There is a positive and significant relationship between the likelihood that someone walks to work, takes public transit, or owns one car and walkability. The magnitudes of the IV coefficients on walkability are greater than the OLS coefficients, suggesting that instrumenting for the walkability index improved the estimation. The OLS and IV results produce similar results in terms of the relationship between the dependent variables and walkability and the significance of the results, although the IV model shows a negative relationship between the likelihood that someone owns one car and walkability while the OLS results show a positive relationship. The IV results also indicate that median income, education, and demographics explain some of the variation in car ownership and use. These coefficients are nearly all the same as the OLS coefficients in terms of the direction of the relationship and the significance of the relationship.

Tables 7 though 10 in the appendix show the IV regressions using each component of the walkability index rather than the entire. These components include meters to the nearest transit stop, intersection density, employment entropy, and housing entropy. For the most part, these components behave similarly to the walkability index as they are almost all significant and have the same directional relationship as the walkability index. However, employment entropy and household entropy are generally less significant than the other components of the walkability index. None of the entropy coefficients are significant at more than the 0.05 level and more than half of the coefficients are significant at the 0.10 level or not at all. Unlike the entropy coefficients, nearly all the coefficients on intersection density and meters to transit are significant at the 0.01 level. However, the entropy components are each weighted as $1/6^{th}$ of the walkability index while the meters to transit and intersection density are both weighted as $1/3^{th}$ of the walkability index. Therefore, meters to transit and intersection density are stronger indicators of walkability according to the index.

Table 3. IV Results								
Variables	First Stage	Drove	Walked	Public Transit	No Car	One Car	Two Cars	Three+ Cars
Homes built before 1960	1.46*** (0.21)	-	-	-	-	-	-	-
Walkability Index	-	-0.35*** (0.07)	0.45*** (0.09)	0.19*** (0.06)	0.41*** (0.07)	-0.14*** (0.06)	-0.10*** (0.04)	-0.22*** (0.06)
Education	0.27*** (0.03)	0.22 (0.53)	-0.39 (0.75)	0.74* (0.44)	-1.06* (0.59)	2.30*** (0.43)	0.35 (0.30)	-1.66*** (0.501
Log of Income	-0.21*** (0.01)	-0.97*** (0.37)	0.43 (0.48)	0.27 (0.31)	-0.67* (0.39)	-1.77*** (0.30)	0.38* (0.23)	1.31*** (0.35)
Black	-0.08*** (0.02)	-1.36*** (0.14)	0.30 (0.21)	2.47*** (0.14)	0.86*** (0.16)	0.10 (0.09)	-0.51*** (0.06)	-0.68*** (0.13)
Hispanic	0.20*** (0.03)	0.13 (0.34)	-1.31*** (0.45)	2.13*** (0.28)	-0.97** (0.41)	1.15*** (0.26)	0.52*** (0.18)	1.16*** (0.34)
Job density	0.09*** (0.01)	0.47*** (0.119)	-0.79*** (0.165)	0.044 (0.109)	-0.51*** (0.130)	0.33** (0.0930)	0.13* (0.0699)	0.14 (0.109)
Constant	-0.09* (0.06)	2.41*** (0.32)	-0.77 (0.51)	-7.27*** (0.41)	0.14 (0.42)	-0.34 (0.24)	-1.83*** (0.14)	-2.82*** (0.23)
Observations	1,209	1,373	1,087	1,257	1,375	1,375	1,375	1,375
R-squared	0.504		0.053	0.600	0.450	0.023	0.532	0.574
Hausman test (p-value)	-	0.00	0.00	0.02	0.00	0.00	0.01	0.00
Basmann test (p-value)	-	0.94	0.04	0.32	0.61	0.24	0.17	0.08

*Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The Hausman test tests the null hypothesis that the walkability index is exogenous. The Basmann test tests the null hypothesis that instruments are uncorrelated with the error term. If we cannot reject the null, instruments are exogenous.*

VI. Quantification

So far, I have discussed the results in terms of the direction of the relationship and the significance of the coefficients, rather than interpreting the magnitude of the coefficients. The log-odds transformation creates some interpretation issues for the coefficients on the log-odds variables in terms of their effect on the probability of driving to work (p). To interpret a coefficient as a change in p, I take the derivative of the log odds with respect to p. A one-unit change in the walkability index causes a β·p(1-p) change in p, the fraction of commuters in each census tract who drive to work.

p = The fraction of households in each census tract that drive to work.

w = National Walkability Index

$$\log \frac{1}{1-p} = \log odds\ dependent\ variable$$

$$\beta = \log odds\ coefficient$$

Taking derivatives and substituting:

$$\beta = \partial \frac{(\log(\frac{p}{1-p}))}{\partial w}$$

$$\partial \log\left(\frac{p}{1-p}\right) = \frac{1}{p} * \frac{1}{1-p} * \ \partial p$$

$$\frac{\partial p}{\partial w} = \beta * p(1 - p)$$

Applying this transformation to the coefficients on walkability for all dependent variables gives us more interpretable coefficients. Table 4 shows the predicted change in the fraction of individuals in each census tract in terms of car ownership in use when the walkability score of a census tract increases by 5%. According to the IV results, the fraction of people driving in a census tract decreases by 0.81 with a 5% increase in walkability and the fraction of people that own no cars increases by 0.52. These results indicate that walkability can have a relatively large impact on commuting behavior. However, the results were calculated using the mean value of p for all urban census tracts in Maryland. This means that a few outliers may be amplifying the effects of walkability for this interpretation of coefficients. That said, the results do show a fairly large impact of walkability on the fraction of people who drive and own cars. Tables 2 and 3 confirm the directional relationship shown in Table 4 and confirm that the coefficients are significant. This table also shows that the OLS and IV results differ quite a bit. Across all dependent variables, the IV results show a much greater impact from the increase in walkability. This confirms that instrumenting for the walkability index is a good idea.

Table 4. Change in a dependent variable associated with increasing the walkability score by one standard deviation (3.45)

Dependent Variables	OLS Results	IV Results
Fraction driving	-0.02	-0.17
Fraction walking	0.01	0.03
Fraction public transit	0.02	0.05
Fraction no car	0.02	0.10
Fraction one car	0.03	-0.10
Fraction two cars	-0.01	-0.08
Fraction three+ cars	-0.03	-0.15

Table four shows the changes in dependent variables associated with increasing a census tract's walkability score (which is measured out of 20) by 1 and by one standard deviation in the distribution of walkability scores, 3.45. For both OLS and IV results, the relationship between owning one car and walkability changes from positive to negative. The table also shows that walkability does influence commuting behavior. Living in a census tract one standard deviation above the mean walkability in urban census tracts decreases the share of people driving to work by 0.17 and increases the share of people with no car by .10. Moreover, it decreases the share of people who own two cars or three or more cars by 0.08 and 0.15 respectively.

VII. Conclusion

This paper presents an investigation into how walkability impacts driving behavior in Maryland. The results find that more walkable census tracts are associated with a decrease in the likelihood that someone drives to work and an increase in the likelihood that they walk to work. Additionally, the results find that there is a significant increase in the likelihood that an individual owns no cars at all and a decrease in the likelihood that an individual owns two or three or more cars. The effects on the likelihood that an individual owns one car are inconclusive. This analysis also shows the importance of controlling for demographic variables, income, education, and job density.

VIII. Appendix

Table 6. Description of Variables

Variable	Description	Source	Observation
Walkability Index (0 to 20)	Weighted score of the sum of intersection density, distance to nearest transit stop, and land use.	EPA Smart Location Database (SLD)	Census Block Group
Meters to transit	Distance from the population-weighted centroid to nearest transit stop (meters).	EPA SLD	Census Block Group
Intersection density	Street intersection density (weighted, auto-oriented intersections eliminated).	EPA SLD	Census Block Group
Land use (employment)	Employment entropy measuring measures the diversity of eight different employment types (office, retail, industrial, service, entertainment, education, health, and public sector) in a block group on a 0 to 1 scale.	EPA SLD	Census Block Group
Land use (HH and emp.)	Housing and employment entropy measuring diversity of employment and housing type.	EPA SLD	Census Block Group
Median income	Log of median income.	American Community Survey (ACS) 5-Year Estimates	Census Tract
Job Density	Jobs within a 45-minute drive of the 2010 population-weighted centroid in a census block group.	EPA SLD	Census Block Group
Black	Fraction of a census tract population that identifies as Black.	ACS 5-Year Estimates	Census Tract
Hispanic	Fraction of a census tract population that identifies as Hispanic.	ACS 5-Year Estimates	Census Tract
Sixty-five plus	Fraction of a census tract that is sixty-five or older.	ACS 5-Year Estimates	Census Tract
Education	Fraction of a census tract that has a bachelor's degree or more.	ACS 5-Year Estimates	Census Tract

Fraction before 1960	Fraction of homes in a census tract that were built before 1960.	ACS 5-Year Estimates	Census Tract
Drove	Fraction of individuals in a census tract that drives to work.	ACS 5-Year Estimates	Census Tract

Table 6. Description of Variables

Variable	Description	Source	Observation
Drove alone	Fraction of individuals in a census tract that drives alone to work.	ACS 5-Year Estimates	Census Tract
Public transit	Fraction of individuals in a census tract that takes public transit to work.	ACS 5-Year Estimates	Census Tract
Carpooled	Fraction of individuals in a census tract that carpools to work.	ACS 5-Year Estimates	Census Tract
Walked	Fraction of individuals in a census tract that walks to work.	ACS 5-Year Estimates	Census Tract
Biked	Fraction of individuals in a census tract that bikes to work.	ACS 5-Year Estimates	Census Tract
Worked from home	Fraction of individuals in a census tract that works from home.	ACS 5-Year Estimates	Census Tract
Commute time	Average commute time in a census tract in minutes.	ACS 5-Year Estimates	Census Tract
No cars	Fraction of individuals in a census tract that own no cars.	ACS 5-Year Estimates	Census Tract
One car	Fraction of individuals in a census tract that own one car.	ACS 5-Year Estimates	Census Tract
Two cars	Fraction of individuals in a census tract that own two cars.	ACS 5-Year Estimates	Census Tract
Three plus cars	Fraction of individuals in a census tract that own three or more cars.	ACS 5-Year Estimates	Census Tract

Table 7. IV Regressions with Meters to Transit

Variables	First Stage	Drove	Walked	Public Transit	No Car	One Car	Two Cars	Three+ Cars
Homes built before 1960	-4,412***							
	(323.5)							
Meters To transit		0.001***	0.003***	0.001***	0.003***	0.001***	0.0006**	0.001***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Education	-468.6***	-0.71***	0.71	0.99***	-0.002	1.58***	-0.13	-2.17***
	(53.97)	(0.27)	(0.46)	(0.28)	(0.37)	(0.28)	(0.19)	(0.38)
Log of Median Income	0.00414***	-0.18	-0.45	-0.20	-1.53***	-1.37***	0.73***	1.73***
	(0.000254)	(0.18)	(0.28)	(0.18)	(0.24)	(0.19)	(0.16)	(0.27)
Black	-47.12	-0.65***	-0.57***	1.83***	0.07	0.15**	-0.35***	-0.25***
	(30.16)	(0.09)	(0.17)	(0.10)	(0.14)	(0.08)	(0.05)	(0.09)
Hispanic	-353.3***	0.19	-1.23***	1.45***	-0.97***	0.63***	0.43***	1.24***
	(62.42)	(0.21)	(0.37)	(0.22)	(0.29)	(0.20)	(0.13)	(0.27)
Job Density	-0.00309***	1.34E-06	-5.69e-6**	5.70e-06**	-2.58e-06*	3.95e-06***	1.09E-07	-1.32E-07
	(0.000125)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	-4,412***	0.774***	-0.338	-3.492***	2.596***	0.262*	-2.434***	-5.484***
	(323.5)	(0.16)	(0.29)	(0.18)	(0.21)	(0.15)	(0.10)	(0.21)
Obs.	1,385	1,207	956	1,142	1,209	1,209	1,209	1,209
R-squared	0.651	0.294	0.043	0.675	0.539	0.185	0.589	0.623

*Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 8. IV Regressions with Intersection Density

Variables	First Stage	Drove	Walked	Public Transit	No Car	One Car	Two Cars	Three+ Cars
Homes built Before 1960	1,424***							
	(139.4)							
Intersection Density		-0.00580**	0.00829***	0.00252**	0.00695***	-0.00198**	-0.00133*	-0.00443**
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Education	133.5***	-0.924***	0.941**	1.155***	0.305	1.433***	-0.222	-2.217***
	(16.35)	(0.22)	(0.38)	(0.27)	(0.43)	(0.24)	(0.17)	(0.39)
Log of Income	-0.0009***	0.131	-0.709***	-0.385**	-1.930***	-1.207***	0.834***	1.866***
	(0.00)	(0.14)	(0.24)	(0.15)	(0.27)	(0.14)	(0.12)	(0.25)
Black	7.311	-0.722***	-0.388***	1.871***	0.16	0.126*	-0.353***	-0.298***
	(6.996)	(0.07)	(0.15)	(0.10)	(0.13)	(0.07)	(0.05)	(0.10)
Hispanic	0.294	-0.653***	-0.132	1.842***	0.0619	0.299*	0.213**	0.673***
	(16.54)	(0.14)	(0.29)	(0.19)	(0.24)	(0.16)	(0.10)	(0.21)
Job density	0.0003***	-2.95e-06***	-3.70E-07	7.73e-06***	2.76e-06***	2.15e-06***	-1.07e-06***	-2.81e-06**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	50.78***	2.949***	-3.537***	-4.438***	-0.0295	1.041***	-1.914***	-3.891***
	(7.275)	(0.28)	(0.46)	(0.31)	(0.51)	(0.25)	(0.23)	(0.47)
Observations	1,385	1,207	956	1,142	1,209	1,209	1,209	1,209
R-squared	0.468	0.489	0.279	0.663	0.562	0.346	0.636	0.623

*Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 9. Household and Employment Entropy

Variables	First Stage	Drove	Walked	Public Transit	No Car	One Car	Two Cars	Three+ Cars
Homes built Before 1960	-0.0635							
	(0.244)							
Employment HH entropy		5.65	-18.66*	-0.63	-5.27	-0.60	-0.15	8.942*
		(3.65)	(10.15)	(2.47)	(4.25)	(2.23)	(1.35)	(5.17)
Education	0.118***	-2.488***	4.362***	1.647***	2.045***	1.141***	-0.442***	-3.871***
	(0.0359)	(0.42)	(1.35)	(0.29)	(0.51)	(0.20)	(0.15)	(0.65)
Log of Income	-6.7e-07***	1.212***	-2.840***	-0.750***	-3.132***	-0.988***	0.996***	2.961***
	(1.66e-07)	(0.23)	(0.69)	(0.15)	(0.29)	(0.12)	(0.09)	(0.37)
Black	-0.154***	0.24	-3.335**	1.739***	-0.74	0.04	-0.370*	1.19
	(0.0194)	(0.61)	(1.66)	(0.42)	(0.73)	(0.39)	(0.22)	(0.87)
Hispanic	-0.0504	-0.34	-0.55	1.762***	-0.25	0.31	0.230**	1.073**
	(0.0400)	(0.29)	(0.87)	(0.21)	(0.35)	(0.21)	(0.11)	(0.44)
Job Density	1.80e-08	-4.75e-06***	0.00	8.50e-06***	4.95e-06***	1.49e-06***	-1.51e-06***	-4.10e-06***
	(7.82e-08)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.564***	-2.17	10.72	-3.360**	5.117*	0.92	-2.159**	-10.74***
	(0.0156)	(2.34)	(6.65)	(1.56)	(2.74)	(1.40)	(0.85)	(3.33)
Observations	1,385	1,207	956	1,142	1,209	1,209	1,209	1,209
R-squared	0.097	0.489	0.279	0.663	0.562	0.346	0.636	0.623

*Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

Table 10. Employment Entropy								
Variables	First Stage	Drove	Walked	Public Transit	No Car	One Car	Two Cars	Three+ Cars
Homes built Before 1960	-0.642***							
	(0.218)							
Employment Entropy		11.74**	-20.31**	-5.754*	-13.23**	3.028	2.126	10.40*
		(5.40)	(9.55)	(3.46)	(6.49)	(2.18)	(1.75)	(5.34)
Education	0.194***	-3.247***	4.579***	2.220***	3.023***	0.741***	-0.697***	-4.172***
	(0.0337)	(0.75)	(1.34)	(0.48)	(0.89)	(0.26)	(0.22)	(0.72)
Median Income	-3.68e-07**	0.738***	-1.556***	-0.611***	-2.667***	-0.988***	0.980***	2.313***
	(1.53e-07)	(0.19)	(0.38)	(0.13)	(0.27)	(0.09)	(0.08)	(0.24)
Black	-0.184***	1.602	-4.508**	0.731	-2.443*	0.724*	0.0662	1.743*
	(0.0192)	(1.07)	(1.98)	(0.68)	(1.28)	(0.44)	(0.34)	(1.04)
Hispanic	0.0544	-1.203**	0.86	2.108***	0.697	0.16	0.114	0.156
	(0.0395)	(0.57)	(1.09)	(0.35)	(0.67)	(0.21)	(0.17)	(0.53)
Job Density	-2.95e-07***	-1.16E-06	-4.60E-06	6.73e-06***	8.66E-07	2.46e-06***	-8.34E-07	-9.77E-07
	(7.62e-08)	0.00	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.577***	-5.021*	10.10*	-0.665	9.033**	-1.134	-3.426***	-10.75***
	(0.0150)	(2.97)	(5.41)	(1.85)	(3.54)	(1.17)	(0.94)	(2.90)
Observations	1,385	1,207	956	1,142	1,209	1,209	1,209	1,209
R-squared	0.253	0.489	0.279	0.663	0.562	0.346	0.636	0.623

*Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1*

IX. Sources

Arango, Mauricio. 2019. "Assessing Walkability through Parking Prices." Accessed December 2, 2021. https://marangoisa.github.io/AssessingWalkabilityThroughParkingPrices.pdf.

"Basic Information about the Built Environment." EPA. Environmental Protection Agency. 0 Accessed December 2, 2021. https://www.epa.gov/smm/basic-information-about-built-environment.

Bento, Antonio M., Maureen L. Cropper, Ahmed Mushfiq Mobarak, and Katja Vinha. 2005. "The Effects of Urban Spatial Structure on Travel Demand in the United States." Review of Economics and Statistics 87 (3): 466–78. https://doi.org/10.1162/0034653054638292.

Bhat, Chandra R., and Jessica Y. Guo. 2007. "A Comprehensive Analysis of Built Environment Characteristics on Household Residential Choice and Auto Ownership Levels." Transportation Research Part B: Methodological 41 (5): 506–26. https://doi.org/10.1016/j.trb.2005.12.005.

Boarnet, Marlon G., and Sharon Sarmiento. "Can Land-Use Policy Affect Travel Behavior? A Study of the Link between Non-Work Travel and Land-Use Characteristics." Urban Studies 35, no. 7 (1998): 1155–69. http://www.jstor.org/stable/43083936.

Brueckner, Jan K. 2000. "Urban Sprawl: Diagnosis and Remedies." International Regional Science Review 23 (2): 160–71. https://doi.org/10.1177/016001700761012710.

Cao, Xinyu (Jason), Patricia L. Mokhtarian, and Susan L. Handy. 2009. "Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings." Transport Reviews 29 (3): 359–95. https://doi.org/10.1080/01441640802539195.

Chen, Cynthia, Hongmian Gong, and Robert Paaswell. 2007. "Role of the Built Environment on Mode Choice Decisions: Additional Evidence on the Impact of Density." Transportation 35 (3): 285–99. https://doi.org/10.1007/s11116-007-9153-5.

Duranton, Gilles, and Diego Puga. 2020. "The Economics of Urban Density." Journal of Economic Perspectives 34 (3): 3–26. https://doi.org/10.1257/jep.34.3.3.

EPA. Environmental Protection Agency. Accessed March 2, 2022. https://www.epa.gov/transportation-air-pollution-and-climate-change/carbon-pollution-transportation#transportation.

US EPA, OA. 2019. "Smart Location Mapping | US EPA." US EPA. April 19, 2019. https://www.epa.gov/smartgrowth/smart-location-mapping.

Frank, Lawrence Douglas, Brian E. Saelens, Ken E. Powell, and James E. Chapman. 2007. "Stepping towards Causation: Do Build Environments or Neighborhood and Travel Preferences Explain Physical Activity, Driving, and Obesity?" Social Science & Medicine 65 (9): 1898–1914. https://doi.org/10.1016/j.socscimed.2007.05.053.

Guo, Yuntao, Srinivas Peeta, and Sekhar Somenahalli. 2015. "The Impact of Walkable Environment on Single-Family Residential Property Values." Journal of Transport and Land Use, September. https://doi.org/10.5198/jtlu.2015.824.

Handy, Susan, Xinyu Cao, and Patricia Mokhtarian. 2005. "Correlation of Causality between the Built Environment and Travel Behavior? Evidence from Northern California." Transportation Research Part D: Transport and Environment 10 (6): 427–44. https://doi.org/10.1016/j.trd.2005.05.002.

Kitamura, Ryuichi, Patricia L. Mokhtarian, and Laura Daidet. 1997. "A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area." Transportation 24 (2): 125–58. https://doi.org/10.1023/a:1017959825565.

Knaap, Gerrit-Jan & Frece, John. (2007). Smart growth in Maryland: Looking forward and looking back. Idaho Law Review. 43.

Maria Kockelman, Kara. 1997. "Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from San Francisco Bay Area." Transportation Research Record: Journal of the Transportation Research Board 1607 (1): 116–25. https://doi.org/10.3141/1607-16.

Mokhtarian, Patricia L., and Xinyu Cao. 2008. "Examining the Impacts of Residential Self-Selection on Travel Behavior: A Focus on Methodologies." Transportation Research Part B: Methodological 42 (3): 204–28. https://doi.org/10.1016/j.trb.2007.07.006.

Næss, Petter. 2015. "Built Environment, Causality, and Travel." Transport Reviews 35 (3): 275–91. https://doi.org/10.1080/01441647.2015.1017751.

Pinjari, Abdul Rawoof, Ram M. Pendyala, Chandra R. Bhat, and Paul A. Waddell. 2007. "Modeling Residential Sorting Effects to Understand the Impact of the Built Environment on Commute Mode Choice." Transportation 34 (5): 557–73. https://doi.org/10.1007/s11116-007-9127-7.

Scheiner, Joachim, and Christian Holz-Rau. 2007. "Travel Mode Choice: Affected by Objective or Subjective Determinants?" Transportation 34 (4): 487–511. https://doi.org/10.1007/s11116-007-9112-1.

"U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." U.S. Energy Information Administration (EIA) - Ap. Accessed December 2, 2021. https://www.eia.gov/environment/emissions/carbon/.

Wells, Nancy M., and Yizhao Yang. 2008. "Neighborhood Design and Walking." American Journal of Preventive Medicine 34 (4): 313–19. https://doi.org/10.1016/j.amepre.2008.01.019.