

Exploring the importance of detailed environment variables in neighborhood commute mode share models

Robert J. Schneider

University of Wisconsin-Milwaukee
rjschnei@uwm.edu

Lingqian Hu

University of Wisconsin-Milwaukee
hul@uwm.edu

Joseph Stefanich

University of Wisconsin-Milwaukee
stefan23@uwm.edu

Abstract: This paper analyzes the relationship between detailed neighborhood environment variables and commute mode share using a dataset drawn from across the United States and includes model validation results. Representing one of the first studies of its kind, we use United States journey-to-work data to explore the following questions: 1) Which detailed environment variables have significant associations with the proportion of people in a neighborhood who take public transit, walk, or bicycle to work? 2) Does adding detailed environment variables to existing, nationally available neighborhood variables improve the predictive accuracy of work commute mode share models? We use a set of 120 randomly selected census tracts to estimate fractional multinomial logit models that predict walk, bicycle, transit, and automobile commute mode shares. The Base Model includes a set of nine significant, nationally available variables identified from a previous analysis of 5,000 tracts. We test 18 additional detailed neighborhood environment variables and identify five variables that have significant associations with commute mode share: sidewalk coverage (positive association with transit and walk), proximity to a rail station (positive association with transit), bicycle facility density (positive association with bicycle), freeway presence (negative association with walk), and mixed land use (positive association with transit, walk, and bicycle). While these detailed environment variables add clarity to our understanding of factors that influence travel behavior, our validation analysis using 50 separate census tracts does not provide conclusive evidence that these variables improve model accuracy. Further studies with larger sample sizes are needed to determine the optimal set of variables to include to predict automobile, transit, pedestrian, and bicycle commute mode shares.

Article history:

Received: November 10, 2016

Received in revised form:

May 14, 2017

Accepted: November 6, 2017

Available online: October 29,
2018

Copyright 2018 Robert J. Schneider, Joseph Stefanich & Lingqian Hu

<http://dx.doi.org/10.5198/jtl.2018.927>

ISSN: 1938-7849 | Licensed under the [Creative Commons Attribution – Noncommercial License 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies. This paper is also published with additional sponsorship from WSTLUR.

1 Introduction

United States transportation policy is shifting from a 20th-Century focus on automobile mobility to a 21st-Century focus on multimodal choices (Shoup & Lang, 2011; Levinson, 2015). For example, the Fixing America's Surface Transportation (FAST) Act supports transit-oriented development projects and pedestrian and bicycle network connectivity (US Congress, 2015) and the Surgeon General issued a Call to Action to Promote Walking and Walkable Communities (US Department of Health and Human Services, 2015). Of the 50 largest cities, 41 have a published goal to increase walking and 47 have a published goal to increase bicycling (Alliance for Walking and Bicycling, 2016). During 2014 and 2015, local voters approved approximately 70% of ballot measures to provide funding for public transit (Center for Transportation Excellence, 2016).

To help guide transportation and land use decisions in this multimodal context, practitioners need analytical tools that consider automobile, public transit, pedestrian, and bicycle travel. Travel demand tools should provide estimates of travel activity at a sufficiently-small geographic scale so that they can reflect the impacts of planning, engineering, design, and other strategies that are implemented at the local level. Recognizing significant associations between neighborhood environment characteristics and small-scale travel behavior, many researchers over the last two decades have incorporated land use, building design, transit accessibility, street network connectivity, roadway design, and pedestrian and bicycle facility variables into demand models.

However, creating modeling tools that incorporate neighborhood environment characteristics is challenging for a variety of reasons. One important practical challenge is optimizing model inputs. On one hand, if the model inputs do not adequately represent the most important determinants of travel decisions, the outputs may not predict travel behavior accurately. On the other hand, if the inputs are too detailed, they may require significant data collection, measurement, and analysis to create, making the model itself impractical for transportation agencies to apply. Further, models that are developed from a single community may not be accurate in other parts of the country.

Considering these practical challenges, this paper explores two related questions using a national-scale dataset: First, what detailed environment variables have significant associations with the share of people in a neighborhood who take public transit, walk, or bicycle to work? Second, does adding these detailed variables to existing, nationally available variables improve the predictive accuracy of work commute mode share models?

The following sections of the paper review the relationship between neighborhood environment variables and mode share, describe the development of neighborhood-level commute mode share models from a national-scale dataset, and compare the accuracy of models based on nationally available variables with models that include additional detailed neighborhood environment variables.

2 The neighborhood environment and mode share

There is a large body of research on how neighborhood environment characteristics are associated with travel behavior (Forsyth, Hearst, Oakes, & Schmitz, 2008; Saelens & Handy, 2008; Ewing & Cervero, 2010; Forsyth & Krizek, 2010; Sallis, Floyd, Rodríguez, & Salens, 2012; Schneider, 2015). After controlling for neighborhood socioeconomic characteristics (e.g., household size, income, automobile availability), these studies generally find that walking, bicycling, and public transit use are associated with a number of factors that can be measured from nationally available data sources (e.g., US Census American Community Survey and Longitudinal Employer-Household Dynamics), including population density, employment density, and employment mix.

The literature also identifies detailed environment variables associated with mode choice. These

have been gathered from local data sources, measured from secondary sources, or collected in the field. Detailed neighborhood environment variables that have been used in pedestrian, bicycle, or transit demand models include:

- Transportation system characteristics: sidewalk coverage (Kuzmyak, Walters, Bradley, & Kockelman, 2014); proximity to multi-use trails (Hankey et al., 2012; Hankey & Lindsey, 2016); proximity to bicycle facilities (Griswold, Medury, & Schneider, 2011; Strauss & Miranda-Moreno, 2013); proximity to transit stops and stations (Miranda-Moreno & Fernandes, 2011; Kuzmyak et al., 2014; Hankey & Lindsey, 2016); street connectivity and intersection density (Miranda-Moreno & Fernandes, 2011; Kuzmyak et al., 2014); roadway functional classification and the presence of traffic signals (Parkin, Wardman, & Page, 2008; Miranda-Moreno & Fernandes, 2011, Schneider, Henry, Mitman, Stonehill, & Koehler, 2012, Hankey et al., 2012); vehicle volumes (Miranda-Moreno & Fernandes, 2011; Grembek et al., 2014); presence of parking entrances (Strauss & Miranda-Moreno, 2013).
- Land use characteristics: proximity to commercial uses, open space, schools, and universities (Griswold et al., 2011; Miranda-Moreno & Fernandes, 2011; Schneider et al., 2012; Strauss & Miranda-Moreno, 2013; Hankey & Lindsey, 2016); land use mix (Strauss & Miranda-Moreno, 2013); distance from the central business district (Hankey et al., 2012).
- Environmental characteristics: slope (Parkin et al., 2008; Griswold et al., 2011; Kuzmyak et al., 2014); and temperature and precipitation (Parkin et al., 2008).

Some detailed variables have been investigated but rarely incorporated into predictive models, including automobile parking cost and availability (Forsyth & Krizek, 2010), parking lot size (Schneider, 2015), roadway aesthetics and lighting (Saelens & Handy, 2008; Sallis et al., 2012), and roadway tree canopy (Schneider, 2015).

A challenge of using detailed neighborhood environment variables for demand estimation is that they are often highly-correlated. For example, many densely-populated neighborhoods have a mix of land uses and excellent pedestrian facilities; many low-density neighborhoods are exclusively residential, have multi-lane roadways, and lack sidewalks and well-designed crosswalks. Factor analysis (Schneider, Shafizadeh, & Handy, 2015) and neighborhood environment indices (Clifton, Singleton, Muhs, & Schneider, 2016a) have been used to address this issue. Another challenge of using neighborhood environment variables in predictive models is self-selection bias. These variables may be associated with transit, pedestrian, or bicycle commuting through residential preferences (e.g., people who like to walk and bicycle may seek to live in neighborhoods where commuting by these modes is already safe and convenient rather than neighborhood environment modifications causing existing residents to change how they commute) (Cao, Mokhtarian, & Handy, 2009). Self-selection bias may lead to either overestimating or underestimating the influence of neighborhood characteristics on mode share (Chatman, 2009).

Due to resource constraints, most studies of detailed neighborhood environment characteristics and travel behavior have been conducted within one or two metropolitan regions. Voulgaris, Taylor, Blumenberg, Brown, and Ralph (2017) recently used nationwide built environment data to identify seven neighborhood types and relate them to individual travel behavior. Our study is different in several ways: first, we estimate the impacts of specific neighborhood variables independently (rather than grouping them into factors); second, we estimate neighborhood-level commute-to-work patterns; third, we define and use secondary sources to measure our own detailed environment variables. Further, we add to the small number of studies that have attempted to test the predictive accuracy of neighborhood-level mode share models (Clifton et al., 2016a, 2016b).

3 Methods

We develop fractional multinomial logit (FMNL) models to predict the proportion of people in a neighborhood who take public transit, walk, or bicycle to work. The FMNL model structure (Papke & Wooldridge, 1996) is useful because it can be used for aggregated data (e.g., proportions of workers in a neighborhood) and accounts for the constraint that mode share proportions must sum to one. This model has been applied in the economics field (Ramalho, Ramalho, & Murteira, 2011; Wang & Wolman, 2014; Mullahy, 2015) but less often in transportation studies (Sivakumar & Bhat, 2002). The model formulation is provided in Section 3.4.

Like previous studies (Cervero, 2002), we compare a Base Model with an Extended Model to determine the value of including additional explanatory variables. Our Base Model uses significant, nationally available explanatory variables that were identified from a previous analysis of 5,000 census tracts, and our Extended Model incorporates additional detailed environment variables. We conducted validation tests on both models to determine their accuracy.

3.1 Sample of neighborhoods

The 73,056 United States census tracts average 4,300 residents and 1,900 workers, so they approximate neighborhood areas. During a previous phase of our study, we used a random process to select 5,000 model development tracts and 1,000 model validation tracts. We excluded tracts that were larger than five square miles (13 square km), tracts with population density less than 500 people per square mile (193 people per square km), tracts with fewer than 200 households or 100 workers, and tracts from Massachusetts (our employment data source did not include information from this state). We also avoided selecting adjacent tracts. More details about this selection process are provided in Schneider, Hu, and Stefanich (2017).

For this phase of the study, we used a random subset of the previously-selected tracts for model development and model validation. We selected 120 of the 5,000 model development tracts and 50 of the 1,000 model validation tracts. Our set of 120 model development tracts was from 34 different states, representing a wide range of geographic areas. More information about these 120 tracts is provided in the following sections.

3.2 Dependent variable

Our dependent variable, commute mode share, is based on American Community Survey (ACS) 2009-2013 journey-to-work data (US Census Bureau ACS, 2015). The ACS asks respondents who worked at least one hour in the previous week to select the mode that they used to travel to work most often during that week. Based on these responses, the US Census Bureau estimates the total number of workers in each tract who commute by automobile, public transit, bicycling, walking, and other modes, as well as work at home. We focused our study on the share of workers who commute using the following modes: automobile (86% of all US workers), public transit (5.0%), walk (2.8%), and bicycle (0.57%). People who work from home (4.3% of all US workers) or commute by other modes (0.86%) were excluded.

The ACS provides a 90% margin of error for its commute mode share estimates. This is particularly important for neighborhood-level studies of commuting since the number of workers who take transit, walk, or bicycle may be small. The median margin of error intervals around the mean estimate for the 120 model development tracts were +/- 13% for automobile, +/- 2.6% for public transit, +/- 2.3% for walking, and +/- 0.94% for bicycle. Despite some uncertainty, the estimates provided by the ACS are mean values and are the best data available for this analysis.

Commute trips are important because approximately one-sixth of United States trips are made to and from work (Federal Highway Administration, 2009). Further, workplaces are a common anchor for all daily travel, so the work commute mode choice often influences the travel mode used on other trips (Krizek, 2003; McGuckin & Srinivasan, 2005). Finally, the American Community Survey (ACS) is the only neighborhood-level travel data source that covers the entire United States, but it only gathers data on work commuting, so our focus on work trips is also driven by data availability.

3.3 Explanatory variables

We considered two groups of explanatory variables for our neighborhood commute mode share model. The first group came from nationally available sources and was refined during an earlier phase of this study. The second group was developed from detailed observations of online maps and aerial imagery.

3.3.1 Nationally available variables

Previously, we developed a neighborhood-level commute mode share model from nationally available variables (Schneider et al., 2017). Initially, we considered 16 neighborhood socioeconomic, four general neighborhood environment, and five regional variables from the ACS, 2011 Longitudinal Employer-Household Dynamics (LEHD) (US Census Bureau LEHD, 2015), and 1981-2010 National Oceanic and Atmospheric Administration (NOAA) climate normals (NOAA, 2014). We used the set of 5,000 randomly selected, non-adjacent census tracts to identify which variables were significantly associated with transit, walk, and bicycle mode shares. Several variables were excluded because of high correlation with other variables (e.g., share of households with no vehicle, share of residents with a college degree) or dropped from the model due to lack of significance (e.g., share of male residents, share of children, wet climate). We also explored a variety of ways to represent certain variables. For example, we measured employment density within 0.5-mile, one-mile, and two-mile (0.8-km, 1.6-km, and 3.2-km) straight-line buffers from the center of each tract. The one-mile buffer fit best in the model. We tested square and square-root values of population and employment density variables to explore non-linear relationships with commute mode shares. However, these alternatives were no better than the linear relationships. The preliminary modeling process identified 19 nationally available variables that were associated with at least one of the commute mode shares at a 95% confidence level (Table 1). We present theoretical reasons for testing specific nationally available variables, the modeling process, and results in more detail elsewhere (Schneider et al., 2017). We started with the 19 nationally available variables that were significant in our previous commute mode share model to develop the Base Model for this study. These variables summarized on the top of Table 2.

3.3.2 Detailed neighborhood environment variables

We used Google Maps (Google, Inc., 2015a) data and Google Earth (Google, Inc., 2015b) aerial imagery to collect 18 additional detailed neighborhood environment variables for our 120 modeling and 50 validation tracts. These detailed variables included pedestrian, bicycle, transit, and automobile facilities as well as land use characteristics. They are summarized on the bottom of Table 2. We chose specific variables that we anticipated would be associated with neighborhood mode shares and were relatively easily to measure or estimate. We used the Google Maps search function to find shopping centers, universities, trails, and transit stops and stations in the vicinity of each census tract. We zoomed to different scales in Google Earth to estimate sidewalk coverage, estimate the size of parking lots, identify the presence of non-residential land uses, measure distances to employment and activity

centers, count roadway lanes, and identify and measure bike facilities. All variables are described in detail in Appendix A.

We expected tracts with more sidewalks to have higher walk mode shares; tracts with more on-road bicycle facilities and nearby trails to have higher bicycle mode shares; and tracts with higher transit stop density and nearby rail stations to have higher transit mode shares. We thought that tracts near freeways, with more six-lane roadways, and more land devoted to off-street parking would have higher automobile mode shares (and less transit, walk, and bicycle commuting). We expected that tracts closer to major employment centers, such as central business districts, local business districts, shopping centers, and university campuses and tracts with a greater mix of land uses (shorter distances to local employment opportunities) would have lower automobile mode shares (and more transit, walk, and bicycle commuting). Finally, we thought that tracts with a greater range of elevations would discourage human-powered bicycle and walk commuting.

Ideally, we would have liked to collect detailed data for more tracts to increase our sample size for modeling. However, measuring all 18 variables took between 30 and 60 minutes per tract (approximately three work weeks, total, for the 170 tracts), so we did not have resources to collect more. The Google Earth aerial imagery was from 2015, slightly later than the ACS commute data (2009-2013), so it is possible that a few tracts experienced built environment changes between these two time periods. Despite these limitations, the detailed built environment variables were fairly easy to collect and were measured objectively from an aerial imagery source with extensive national coverage.

Table 1: Nationally available variables identified in previous study¹

	Transit	Walk	Bicycle
Socioeconomic Characteristics	+ Elderly (>64) % + Black % + Asian % + Hispanic % - Disabled % + Median income + Unemployed % + Students in College % - Manufacturing %	- Elderly (>64) % - Black % - Hispanic % - Median income + Unemployed % + Students in College % - Construction % - Manufacturing % - Born in same state %	- Black % - Asian % + Students in College % - Manufacturing % - Born in same state %
General Neighborhood Environment Characteristics	+ Population density + Jobs within 1 mi + Pre-1940 housing % + Rental housing %	+ Population density + Jobs within 1 mi + Pre-1940 housing % + Rental housing %	+ Population density + Jobs within 1 mi + Pre-1940 housing % + Rental housing %
Regional Characteristics	+ Northeast state - Southern state - West Coast state	+ Northeast state - Southern state + Warm climate	- Northeast state - Southern state + West Coast state

¹ The previous study is Schneider et al. (2017). Development of a neighborhood commute mode share model using nationally available data. *Transportation*. Retrieved from <https://doi.org/10.1007/s11116-017-9813-z>, 2017.

The previous study identified variables significantly-associated with commute mode shares for 5,000 census tracts.

+ indicates that the parameter sign is positive and $p < 0.05$.

- indicates that the parameter sign is negative and $p < 0.05$.

Automobile mode share is the base. Signs indicate increase or decrease relative to automobile mode share.

Table 2: Variable descriptive statistics

	Model Development		Model Validation	
Number of Census Tracts	120		50	
Dependent Variables¹	Mean	Std Dev	Mean	Std Dev
Automobile share	.8522	.1975	.8789	.1721
Public transit share	.0893	.1473	.0814	.1469
Walk share	.0478	.0866	.0313	.0604
Bicycle share	.0107	.0218	.0085	.0149
Nationally Available Explanatory Variables	Mean	Std Dev	Mean	Std Dev
<i>Socioeconomic Characteristics</i>				
Elderly (>64 years) share	.142	.095	.129	.074
Non-Hispanic Black share	.155	.211	.189	.272
Asian share	.056	.093	.054	.098
Hispanic share	.187	.202	.236	.264
Disabled share	.122	.058	.125	.046
Median Income (thousands of 2013 dollars)	57.830	29.748	54.115	29.136
Unemployed share	.104	.066	.121	.071
In College share	.320	.159	.264	.138
Work in Construction share	.053	.040	.054	.036
Work in Manufacturing share	.094	.058	.095	.072
Born in Same State share	.554	.171	.586	.187
<i>General Neighborhood Characteristics</i>				
Population per sq. mi. (000s)	8.810	19.145	7.042	9.845
Employment within 1 mi. (000s)	14.331	44.216	6.151	10.310
Pre-1940 housing share	.163	.226	.153	.183
Rental housing share	.442	.249	.436	.245
<i>Regional Characteristics</i>				
Warm climate (0 = no, 1 = yes)	.292	.456	.280	.454
Northeast state (0 = no, 1 = yes)	.175	.382	.220	.419
Southern state (0 = no, 1 = yes)	.300	.460	.240	.431
West Coast state (0 = no, 1 = yes)	.208	.408	.200	.404

¹. The shares reported for automobile, public transit, walk, and bicycle commuting are based on the total number of workers who commute by these 4 modes. Workers who work from home or commute by other means are excluded.

Table 2: Variable descriptive statistics (continued)

	Model Development		Model Validation	
Number of Census Tracts	120		50	
Detailed Environment Explanatory Variables²	Mean	Std Dev	Mean	Std Dev
<i>Transportation System Characteristics</i>				
Proportion of roadways with sidewalks	.649	.362	.708	.356
Bicycle facility density (mi. per sq. mi.)	2.390	4.008	2.562	3.770
Major trail between tract & local business district (0 = no, 1 = yes)	.175	.382		
Major trail between tract & closest university campus (0 = no, 1 = yes)	.200	.402		
Transit stops per sq. mi.	21.698	31.017		
Rail station within 0.5 mi. of tract (0 = no, 1 = yes)	.258	.440	.260	.443
Proportion of developed land used for off-street parking	.132	.158		
Proportion of roadways with 2 or 3 lanes	.834	.146		
Proportion of roadways with 6 or more lanes	.040	.081		
Freeway within the tract (0 = no, 1 = yes)	.483	.502	.400	.495
Freeway between tract & employment ctr. (0 = no, 1 = yes)	.092	.290		
<i>Land Use Characteristics</i>				
Local business district near tract (0 = no, 1 = yes)	.583	.495		
Distance to regional central business district (mi.)	12.290	13.691		
Distance to closest university campus (mi.)	6.397	8.493		
Distance to closest shopping center (mi.)	1.626	4.347		
Tract has a mix of land uses (0 = no, 1 = yes)	.575	.496	.460	.504
Range of elevations at edge of tract (ft.)	118.583	142.593		

² All detailed neighborhood environment variables were collected for the 120 model estimation tracts, but only the five detailed environment variables included in the model were collected for the 50 validation tracts.

Note: 1 sq. mi. = 2.59 sq. km; 1 mi. = 1.6 km; 1 ft. = 0.305 m

3.4 Fractional multinomial logit model

We estimated mode share proportions for automobile, public transit, walk, and bicycle commuting using the FMNL model. This model has a structure similar to a standard multinomial logit model, but it is applied to aggregated data and the dependent variable proportions are constrained to sum to one. Formally, the model considers a set of n independent observations $\{y_{ij}, \mathbf{X}_i\}$ for census tract $i = 1, 2, \dots, n$ where y_{ij} is the dependent variable representing the proportion of commuters using mode j out of the four alternatives $J = \{\text{automobile, public transit, walk, or bicycle}\}$ and \mathbf{X}_i is a vector of explanatory

variables taken from Table 2. It assumes $0 \leq y_{ij} \leq 1$ and that $\sum_{j=1}^J y_{ij} = 1$.

FMNL uses the logistic function so that the expected value of y_{ij} for a particular mode j is expressed as

$$E(y_{ij} | \mathbf{X}_i) = \frac{\exp(\beta_j \mathbf{X}_i)}{\sum_{j=1}^J \exp(\beta_j \mathbf{X}_i)} \quad (1)$$

where β_j is the vector of parameters corresponding with each explanatory variable in \mathbf{X}_i . More detail about the model formulation is provided by Papke and Wooldridge (1996).

3.5 Modeling process

We estimated a series of FMNL neighborhood commute mode share models using our 120-tract model development database. First, we developed a Base Model using only nationally available variables. We started by considering all 19 significant variables from our previous commute mode share model that was based on 5,000 tracts. Likely due to the smaller sample size in this study, some of these variables were statistically insignificant. After removing insignificant variables and re-estimating the model, our Base Model includes nine variables with statistically-significant parameters for at least one of the three modes (95% confidence level).

Second, we developed an Extended Model that supplemented the nine nationally available variables with our set of detailed environment variables. We initially considered all 18 detailed environment variables, but some were not significant for any mode. We removed variables with the least accurate parameter estimates first and re-estimated a series of models until we arrived at a model that only included statistically-significant variables. Our Extended Model includes the nine nationally available variables and five detailed environment variables.

During the modeling process, we checked for correlation between explanatory variables. The Extended Model includes one pair of variables with moderate correlation: bicycle facility density and the number of jobs within one mile (1.6 km) of the center of the tract ($r = 0.65$). Both of these variables are theoretically important and significant in the model. The largest variance inflation factor among all 14 variables used in the Extended Model was 2.2, which does not suggest a problem with multi-collinearity.

We used Stata statistical software (StataCorp, 2007) to estimate the FMNL model parameters. The models used automobile as the base mode, so explanatory variable parameters represent the attractiveness of public transit, walk, and bicycle modes relative to automobile commuting. We assessed the accuracy of the Base Model and Extended Model using the separate set of 50 validation census tracts.

4 Results

The first part of this section summarizes the FMNL neighborhood commute mode share models. The second part summarizes validation results, comparing the accuracy of the Base Model and Extended Model.

4.1 FMNL neighborhood commute models

The Base Model and Extended Model are presented in Table 3. All nine nationally available variables have statistically-significant parameters (95% confidence level) for at least one mode in both models. Transit commuting is positively associated with census tract shares of elderly, non-Hispanic Black, and Hispanic residents. It is also positively associated with jobs near the tract, pre-1940 housing, being in a Northeast state, and rental housing (Base Model only). Walk commuting is positively associated with the share of tract students in college, jobs near the tract, rental housing, and being in a Northeast state. Bicycle commuting is positively associated with the share of tract workers in manufacturing, pre-1940 housing, rental housing, and not being in a Northeast state. Bicycle commuting is negatively associated with the number of jobs within one mile (1.6 km) of the center of the tract (Extended Model only). Comparing the overall fit of the Base Model and Extended Model, the five additional detailed environment variables result in only a slight increase in log pseudo likelihood (-50.98 to -49.57). Yet, the Ex-

tended Model suggests that several detailed environment variables can add depth to our understanding of factors associated with neighborhood mode shares. After controlling for the other nationally available variables:

- Transit commuting is positively associated with having a rail station within 0.5 mi (0.8 km) of the tract, a mix of land uses, and more complete sidewalk coverage. These variables underscore the importance of providing pedestrian access to transit stations and developing mixed-use environments that provide daily activity locations within close proximity to support taking public transit to work.
- Walk commuting is positively associated with mixed land uses and sidewalk coverage. These variables represent convenience—short distances to activity locations—and safety—infrastructure to support walking along roadways. We also found that walk commuting was negatively associated with the presence of a freeway in the tract, suggesting that major automobile infrastructure is a barrier to pedestrians.
- Bicycle commuting is positively associated with bicycle facility density and land use mix. Similar to walking, the convenience of short distances to a variety of activity locations and the perceived safety of bicycle infrastructure appear to contribute to higher neighborhood bicycle commute mode shares.

In general, adding the five detailed environment variables to the Extended Model had minimal influence on the nationally available variable parameter estimates. However, the associations between transit commuting and tract shares of pre-1940 housing and rental housing were weaker in the Extended Model than the Base Model. In the Base Model, pre-1940 housing and rental housing may have picked up some of the influence of having mixed land use, sidewalks, or being close to a rail station on transit mode share. Also, the number of jobs within one mile (1.6 km) of the center of the tract was not associated with bicycle commuting in the Base Model, but it had a negative association with bicycling in the Extended Model. One explanation for this finding is that adding the bicycle facility density and land use mix variables in the Extended Model provided a more complete representation of factors related to bicycle commuting, allowing the job proximity parameter to become more accurate. In fact, the negative sign on this parameter estimate supports the nuanced view that tracts with more jobs located within one mile have lower shares of bicycle commuting, possibly due to people choosing to walk short distances to jobs.

Table 3: Fractional Multinomial Logit Models of Census Tract Commute Mode Shares

	Base Model (coefficients)			Extended Model (coefficients)		
	Transit	Walk	Bicycle	Transit	Walk	Bicycle
<i>Constant</i>	-4.552**	-6.133**	-5.533**	-5.161**	-6.500**	-6.172**
<i>Nationally Available Variables</i>						
Elderly (>64 years) share	3.438**	1.860	2.491*	2.624*	1.029	1.893
Non-Hispanic Black share	1.890**	.118	-1.516	2.180**	.014	-.896
Hispanic share	1.653**	-.019	-1.479	1.213**	-.233	2.089
In College share	-.498	2.604**	-.433	-.423	2.244**	.170
Work in Manufacturing share	-5.681**	-.455	-7.058*	-5.331*	-.115	-6.184*
Employment within 1 mi. (000s)	.0141**	.0130**	-.00314	.00898**	.00995**	-.0150**
Pre-1940 housing share	1.700**	1.374*	2.777**	.861*	.879	2.568**
Rental housing share	1.338**	2.277**	3.197**	.649	2.332**	2.903**
Northeast state (0 = no, 1 = yes)	1.287**	1.188**	-1.078*	1.301**	1.280**	-.976*
<i>Detailed Environment Variables</i>						
Proportion of roadways with sidewalks				.776*	.792*	-.526
Bicycle facility density (mi. per sq. mi.)				.00166	.023	.191**
Rail station within 0.5 mi. of tract (0 = no, 1 = yes)				.928**	-.118	-.110
Freeway within the tract (0 = no, 1 = yes)				.114	-.685**	-.329
Tract has a mix of land uses (0 = no, 1 = yes)				.531**	.704**	1.074**
Log pseudo likelihood		-50.977			-49.568	
Sample size		120			120	

Note: * = significant at $p < 0.05$; ** = significant at $p < 0.01$

4.2 Model validation

We compared the accuracy of the Base Model and Extended Model using our set of 50 validation census tracts. To make our analysis comparable between modes with very large shares (automobile) and small shares (bicycle), we tested whether or not the predicted mode shares were within successive standardized error ranges around the actual mode share values (Table 4). In order to see how many tracts had predicted values within a narrow range of accuracy, we set our first error interval using a z-score of 0.063, which is associated with the 5% confidence level for a two-tailed distribution. To standardize the range of accuracy for each mode, we multiplied the z-score by the standard deviation of each mode share for the 50 tracts (0.172 for automobile, 0.147 for public transit, 0.0604 for walk, and 0.0149 for bicycle). The standardized ranges for this first error level were +/-1.08% for automobile, +/-0.92% for public transit, +/-0.38% for walk, and +/-0.09% for bicycle. For example, we found that 12% of tracts had Base Model automobile mode share estimates within this narrow interval of the actual automobile mode share. We calculated standardized ranges for successively larger error intervals and did a similar analysis at each.

As shown by the bold percentages in Table 4, the Extended Model estimates of automobile, public transit, and bicycle mode shares were more accurate than the Base Model for most of the ten error intervals. However, the Base Model had more accurate estimates at some of the largest error intervals for automobile and public transit mode shares. It is unlikely that the largest error intervals would be acceptable for use in practice, so we concluded that the Extended Model performed better for the automobile, public transit, and bicycle modes. In contrast, the Base Model was slightly more accurate than

the Extended Model for nearly all error intervals for predicting pedestrian mode share.

Many of the incorrect estimates are likely due to zero or near-zero mode shares, which are difficult to estimate from a FMNL model. For example, even if no workers commuted by walking, the tract will include variables that are at least somewhat related to walking, so the FMNL model will estimate some walk more share. This is likely to have affected our results since 43 of the 50 validation census tracts have less than 1% mode share for at least one mode.

Table 4: Validation Results

	<i>Base Model</i>							
	Auto		Public Transit		Walk		Bicycle	
Error Level (z-score)	Interval around mean value	Percent Within Interval	Interval around mean value	Percent Within Interval	Interval around mean value	Percent Within Interval	Interval around mean value	Percent Within Interval
1 (0.063)	+/-1.08%	12%	+/-0.92%	18%	+/-0.38%	14%	+/-0.09%	8%
2 (0.126)	+/-2.16%	26%	+/-1.85%	28%	+/-0.76%	30%	+/-0.19%	14%
3 (0.189)	+/-3.26%	40%	+/-2.78%	40%	+/-1.14%	48%	+/-0.28%	22%
4 (0.253)	+/-4.36%	48%	+/-3.72%	48%	+/-1.53%	54%	+/-0.38%	32%
5 (0.319)	+/-5.49%	54%	+/-4.68%	52%	+/-1.93%	66%	+/-0.48%	36%
6 (0.385)	+/-6.63%	60%	+/-5.66%	60%	+/-2.33%	72%	+/-0.57%	44%
7 (0.454)	+/-7.81%	66%	+/-6.67%	66%	+/-2.74%	76%	+/-0.68%	50%
8 (0.524)	+/-9.03%	74%	+/-7.70%	72%	+/-3.17%	80%	+/-0.78%	62%
9 (0.598)	+/-10.29%	76%	+/-8.78%	74%	+/-3.61%	82%	+/-0.89%	62%
10 (0.674)	+/-11.61%	82%	+/-9.91%	82%	+/-4.07%	86%	+/-1.01%	70%
	<i>Extended Model</i>							
	Auto		Public Transit		Walk		Bicycle	
Error Level (z-score)	Interval around mean value	Percent Within Interval	Interval around mean value	Percent Within Interval	Interval around mean value	Percent Within Interval	Interval around mean value	Percent Within Interval
1 (0.063)	+/-1.08%	14%	+/-0.92%	20%	+/-0.38%	12%	+/-0.09%	10%
2 (0.126)	+/-2.16%	30%	+/-1.85%	32%	+/-0.76%	30%	+/-0.19%	20%
3 (0.189)	+/-3.26%	46%	+/-2.78%	52%	+/-1.14%	48%	+/-0.28%	30%
4 (0.253)	+/-4.36%	54%	+/-3.72%	54%	+/-1.53%	50%	+/-0.38%	36%
5 (0.319)	+/-5.49%	60%	+/-4.68%	68%	+/-1.93%	60%	+/-0.48%	42%
6 (0.385)	+/-6.63%	64%	+/-5.66%	70%	+/-2.33%	68%	+/-0.57%	50%
7 (0.454)	+/-7.81%	68%	+/-6.67%	74%	+/-2.74%	72%	+/-0.68%	54%
8 (0.524)	+/-9.03%	68%	+/-7.70%	80%	+/-3.17%	76%	+/-0.78%	62%
9 (0.598)	+/-10.29%	68%	+/-8.78%	80%	+/-3.61%	80%	+/-0.89%	64%
10 (0.674)	+/-11.61%	78%	+/-9.91%	80%	+/-4.07%	84%	+/-1.01%	72%

Note: Bold indicates the higher percentage between the Base Model and Extended Model (neither value is bolded for a tie).

5 Considerations and Future Research

We identified several detailed environment variables that are associated with commute mode choice, improving our understanding of travel behavior. Yet, the validation results suggested that our Extended

Model may not be ready for practical application. In addition, the models being tested were based on cross-sectional data, so they do not necessarily represent causal relationships between the explanatory and dependent variables. More work is needed to understand interrelationships between the model input variables. Additional studies could use factor analysis (e.g., Voulgaris et al. 2017) and explore time-series to ensure related sets of model input variables change together appropriately. It may soon be possible to address this issue by collecting detailed explanatory variables from high-resolution aerial imagery taken over multiple years and compare these variables to mode shares from different years of the ACS.

Policy makers should also interpret the results cautiously. Variables in the Extended Model may capture the effects of other unmeasured variables. For example, the variables indicating Northeast, Southern, or West Coast states may represent cultural differences in attitudes towards different modes, broad differences in metropolitan spatial patterns, or differences in humidity and precipitation that affect mode commute mode shares. Further, modifying a particular neighborhood variable (e.g., add bicycle lanes; add sidewalks) may not have the impact on mode share that is expected from the model because of potential self-selection bias (Cao, Mokhtarian & Handy 2009; Chatman 2009). Additionally, the models presented here are based on ACS neighborhood commute mode shares with a margin of error, so the model-predicted mode shares also have this underlying variance. Future extensions of these models could quantify the margin of error around their predicted mode shares.

The FMNL model has similar limitations in predicting mode shares as standard multinomial logit models. In reality, increasing commuting by a particular mode may not draw from all other modes equally. For example, a policy to increase in public transit commuting may realistically result in a corresponding decrease in driving, but the model would draw proportional shares from driving as well as walk and bicycle commuting.

Like similar studies, we were constrained by the amount of time required to collect detailed neighborhood environment variables, so we were only able to analyze 170 census tracts. With only 120 tracts available for model estimation, we were not able to develop a model that incorporated all 37 variables that we expected to have a theoretical relationship with neighborhood commute mode share. We have several suggestions to improve on our analysis.

- *Increase the size of the model development dataset.* Increasing the sample size could identify additional significant variables. To explore this possibility, we also estimated our Extended Model using all 170 census tracts (combined the validation tracts with the modeling dataset). Yet, this model was similar to the Extended Model, as only three new parameters became statistically significant (the percentage of housing built before 1940 was positively related to walk commuting, the percentage of students in college was negatively related to transit commuting, and sidewalk coverage was positively related to transit commuting) while two parameters lost significance (sidewalk coverage was no longer related to walk commuting, and mixed land use was no longer related to transit commuting). Even larger samples may be necessary to test additional theoretically-important relationships.
- *Use stratified random sampling.* Even if data collection resources limit researchers to a small dataset, it may be possible to improve model performance using stratified random sampling. Rather than selecting tracts randomly, which tends to include many tracts with small walk, bicycle, and transit commute mode shares, the tracts could be selected from strata with high, medium, and low levels of commuting by each mode. This would help provide more variation in commute mode shares within the dataset, making it easier to identify characteristics that explain this variation.
- *Select tracts that meet minimum thresholds for transit, walk, and bicycle commuting.* Small mode shares are difficult to estimate using the FMNL model, especially with a model developed from a relatively small number of 120 census tracts. This is a particular challenge for bicycle commuting. It may be beneficial to first develop FMNL models for neighborhoods that already

meet minimum thresholds for transit, walk, and bicycle commuting.

- *Increase the size of the validation dataset, and try random hold-out validation.* Further, model accuracy assessments from the 50-tract validation dataset should be viewed with caution. Just a few validation tracts with unusual commuting characteristics could make a notable difference in model accuracy results. Future studies should consider using larger validation sample sizes or repeatedly holding out a different randomly-selected set of validation tracts from all available tracts to generate validation statistics.
- *Test additional explanatory variables.* This study considered 18 detailed neighborhood environment variables, but many other variables could be measured and tested. Automobile parking prices, automobile parking availability, transit service frequency, building setbacks, and motor vehicle speeds could increase the accuracy of the neighborhood mode share models. Future models should also consider other aspects of the natural environment (e.g., proximity to trees and parks; views of water bodies) and social environment (e.g., attitudes toward different travel modes; perceptions of neighborhood traffic safety and crime).

6 Conclusion

We analyzed the relationship between detailed neighborhood environment variables and commute mode share using a sample drawn from across the United States. This involved comparing the accuracy of a Base Model that included only nationally available variables and an Extended Model that also included detailed neighborhood variables. Despite a small sample size of 120 census tracts, our results provide useful information to researchers seeking to integrate neighborhood characteristics into multimodal demand models. Our Extended Model showed that sidewalk coverage, bicycle facility density, proximity to a rail station, freeway presence, and land use mix were associated with neighborhood commute mode share. These variables add clarity to our understanding of factors that influence travel behavior. However, our validation analysis did not provide conclusive evidence that adding detailed environment variables helped improve the predictive accuracy of commute mode share models. The Extended Model generally performed better than the Base Model for predicting automobile, public transit, and bicycle mode shares, but it performed worse for predicting walk mode shares. From a practical perspective, we found only modest support for investing additional time and resources to supplement nationally available variables with detailed environmental variables to improve neighborhood mode share estimates. Further studies with larger sample sizes are needed to determine the optimal set of variables to include to predict automobile, transit, pedestrian, and bicycle commute mode shares.

Acknowledgements

This study was funded by the University of Wisconsin-Milwaukee Office of Research through the Research and Creative Activities Support (RACAS) program award AAA3537. We would like to thank the Office of Research staff for their support and the RACAS proposal reviewers for their helpful input.

References

- Alliance for Biking and Walking. (2016). *Bicycling and walking in the United States: 2016 benchmarking report*. Retrieved from http://www.bikewalkalliance.org/storage/documents/reports/2016benchmarkingreport_web.pdf, 2016.
- Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behavior: A focus on empirical findings. *Transport Reviews*, 29(3), 359–395.
- Center for Transportation Excellence. (2016). Transportation ballot measures, election monitoring database. Retrieved from <http://www.cte.org/elections/past>
- Cervero, R. (2002). Built Environments and mode choice: Toward a normative framework. *Transportation Research Part D*, 7, 265–284.
- Chatman, D. G. (2009). Residential choice, the built environment, and nonwork travel: Evidence using new data and methods. *Environment and Planning A*, 4, 1072–1089.
- Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016a). Representing pedestrian activity in travel demand models: Framework and application. *Journal of Transport Geography*, 52, 111–122.
- Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016b). Development of destination choice models for pedestrian travel. *Transportation Research Part A*, (94) 255–26.
- Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American Planning Association*, (76)3, 265–294.
- Federal Highway Administration. (2009). *National household travel survey*. Washington, DC: Federal Highway Administration.
- Forsyth, A., Hearst, M., Oakes, J. M., & Schmitz, K. H. (2008). Design and destinations: Factors influencing walking and total physical activity. *Urban Studies*, 45(9), 1973–1996.
- Forsyth, A., & Krizek, K. J. (2010). Promoting walking and bicycling: Assessing the evidence to assist planners. *Built Environment*, 36(4), 429–446.
- Google, Inc. (2015a) Google Maps. Retrieved from <https://www.google.com/maps>
- Google, Inc. (2015b) Google Earth, Version 7.1.2.2041. Retrieved from <https://www.google.com/earth/desktop/>
- Grembek, O., Bosman, C., Bigham, J. M, Fine, S., Griswold, J. B., Medury, A., ... Ragland, D. R. (2014). *Pedestrian safety improvement program (Final technical report)*. Berkeley, CA: University of California Berkeley Safe Transportation Research and Education , California Department of Transportation.
- Griswold, J. B., Medury, A., & Schneider, R. J. (2011). Pilot models for estimating bicycle intersection volumes. *Transportation Research Record: Journal of the Transportation Research Board*, 2247, 1–7.
- Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., & Xu, Z. (2012). Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN. *Landscape and Urban Planning*, 107, 307–316.
- Hankey, S., & Lindsey, G. (2016). Facility-demand models of peak-period pedestrian and bicycle traffic: A comparison of fully-specified and reduced-form models. *Transportation Research Record: Journal of the Transportation Research Board*, 2586, 48–58.
- Krizek, K. J. (2003). Neighborhood services, trip purpose, and tour-based travel. *Transportation*, 30, 387–410.
- Kuzmyak, J. R., Walters, J., Bradley, M, & Kockelman, K. M. (2014). *Estimating bicycling and walking for planning and project development: A guidebook*, National Cooperative Highway Research Program, NCHRP report 770. Retrieved from http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_770.

- pdf
- Levinson, D. (2015). *Modernizing American transportation policy, conservative reform network, first edition*. Retrieved from <http://conservativereform.com/wp-content/uploads/2015/11/Transportation3.pdf>, 2015.
- McGuckin, N., & Srinivasan, N. (2005). *The journey-to-work in the context of daily travel: For the Census Data for Transportation Planning Conference*, paper presented at the Census Data for Transportation Planning Conference, May 11–13, Irvine, CA. Retrieved from <http://onlinepubs.trb.org/onlinepubs/archive/conferences/2005/censusdata/resource-journey-to-work.pdf>
- Miranda-Moreno, L. F., & Fernandes, D. (2011). Pedestrian activity modeling at signalized intersections: Land use, urban form, weather and spatiotemporal patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 2264, 74–82.
- Mullahy, J. (2015). Multivariate fractional regression estimation of econometric share models. *Journal of Econometric Methods*, 4(1) 71–100.
- Papke, L. E., & Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6), 619–632.
- Parkin, J., Wardman, M., & Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation*, 35(1) 93–109.
- Ramalho, E. A., Ramalho, J. J., & Murteira, J. M. (2011). Alternative estimating and testing empirical strategies for fractional regression models. *Journal of Economic Surveys*, 25(1) 19–68.
- Saelens, B. E., & Handy, S. L. (2008). Built environment correlates of walking: A review. *Medicine and Science in Sports and Exercise*, 40(7 supplement 1), S550–S556.
- Sallis, J. F., Floyd, M. F., Rodríguez, D. A., & Saelens, B. E. (2012). Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*, 125(5), 729–737.
- Schneider, R. J., Henry, T., Mitman, M. E., Stonehill, L., & Koehler, J. (2012). Development and application of the San Francisco pedestrian intersection volume model. *Transportation Research Record: Journal of the Transportation Research Board*, 2299, 65–78.
- Schneider, R. J. (2015). Local environment characteristics associated with walking and taking transit to shopping districts. *Journal of Transport and Land Use*, 8(2) 125–147.
- Schneider, R. J., Shafizadeh, K., & Handy, S. L. (2015). Method to adjust Institute of Transportation Engineers vehicle trip-generation estimates in smart-growth areas. *Journal of Transport and Land Use*, 8(1), 69–83.
- Schneider, R. J., Hu, L., & Stefanich, J. (2017). Development of a neighborhood commute mode share model using nationally available data. *Transportation*. doi.org/10.1007/s11116-017-9813-z
- Shoup, L., & Lang, M. (2011). *Transportation 101: An introduction to federal transportation policy, Transportation for America*. Retrieved from <http://t4america.org/docs/Transportation%20101.pdf>
- Sivakumar, A., & Bhat, C. (2002). Fractional split-distribution model for statewide commodity-flow analysis. *Transportation Research Record: Journal of the Transportation Research Board*, 1790, 80–88.
- StataCorp. (2007). *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP.
- Strauss, J., & Miranda-Moreno, L. F. (2013). Spatial modeling of bicycle activity at signalized intersections. *Journal of Transport and Land Use*, 6(2) 47–58.
- United States Census Bureau. (2015). American Community Survey (ACS), methodology webpage. Retrieved from <https://www.census.gov/programs-surveys/acs/methodology.html>
- United States Census Bureau. (2015). Longitudinal Employer-Household Dynamics (LEHD). Data webpage. Retrieved from <http://lehd.ces.census.gov/data/>
- United States Congress. (2015). *Fixing America's surface transportation act*. Retrieved from <https://www.congress.gov/bill/114th-congress/house-bill/22/text>

- United States Department of Health and Human Services. (2015). *Step it up! The surgeon general's call to action to promote walking and walkable communities*. Retrieved from <http://www.surgeongeneral.gov/library/calls/walking-and-walkable-communities/call-to-action-walking-and-walkable-communities.pdf>
- United States National Oceanic and Atmospheric Administration (NOAA). (2014). National Climatic Data Center, 1981-2010 U.S. climate normals. Retrieved from <http://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/climate-normals/1981-2010-normals-data>.
- Voulgaris, C. T., Taylor, B. D., Blumenberg, E., Brown, A., & Ralph, K. (2017). Synergistic neighborhood relationships with travel behavior: An analysis of travel in 30,000 US neighborhoods *Journal of Transport and Land Use*, 10(1), 437–461.
- Wang, Z., & Wolman, A. L. (2014). *Payment choice and the future of currency: Insights from two billion retail transactions*, white paper. Richmond, VA: Federal Reserve Bank of Richmond, Research Department.

Appendix

Appendix A is available online at <https://www.jtlu.org/index.php/jtlu/article/view/927>