TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

Research Article

Transportation Research Record
1–11

DOI: 10.1177/0361198119837153
journals.sagepub.com/home/trr
SAGE

# Propensity Score Weighting with Generalized Boosted Models to Explore the Effects of the Built Environment and Residential Self-Selection on Travel Behavior

**Yiling Deng[1] and Yadan Yan[2]**

## Abstract
Many studies have examined the association between the built environment, residential self-selection, and travel behavior. However, few studies have quantified the relative contribution of the built environment itself. Using the 2012 Nanjing Household Travel Survey data, this study applied hierarchical clustering and propensity score weighting to study the effects of the built environment and residential self-selection on travel behavior. First, residents' household locations were classified into four built environment patterns using hierarchical clustering based on six built environment variables by loosely following the "5Ds" (i.e., density, diversity, design, destination accessibility, and distance to transit). Second, a powerful machine learning method, generalized boosted model (GBM), was employed to obtain propensity scores. Propensity score weighting, which is more effective for multiple treatments than matching or stratification, was used to control for residential self-selection. Lastly, the observed effect (OBE), the average treatment effect on the population (ATE), and the built environment proportion (BEP) were calculated for the walking trip frequency, bicycle trip frequency, public transit trip frequency, and vehicle kilometers traveled (VKT) of six pairs of built environment patterns. The results show that a high-density, mixed-use, walkable, and transit-accessible built environment is associated with more walking trips and lower VKT but has no impact on bicycle trips and has an inconsistent impact on public transit trips. The effects of some built environment variables on bicycle and public transit trips are tangled. The residential self-selection effect has the greatest impact on VKT (BEP: 48%–77%), followed by the walking trip frequency (BEP: 62%–74%) and the public transit frequency (BEP: 90%–107%).

With rapid urbanization, problems such as traffic congestion and air pollution have become increasingly serious in urban areas. Increasingly, governments want to implement urban development policies such as New Urbanism and Smart Growth to stimulate meaningful changes in travel behavior—increase walking, biking, and public transit use, and reduce car use. Correctly understanding the correlation of the built environment and travel behavior will provide supporting evidence for the adoption of policies that aim to modify the built environment to achieve integration of transportation and land use.

The influence of the built environment on travel behavior is easily confounded by residential self-selection. Only by excluding the effect of self-selection can policymakers understand the true benefits of built environment-related planning and policy interventions. Cao et al. reviewed 38 empirical studies using different approaches to control for residential self-selection and concluded that the built environment usually has a more significant effect and that residential self-selection also plays an important role in travel behavior (*1*). Given the coexistence of both effects, it is important to know the proportion of the influence of the built environment itself in the influence of the built environment on travel behavior. However, fewer studies have quantified the proportion of the true built environment effect in the total effect (BEP).

[1]College of Civil Engineering and Architecture, Zhejiang University of Technology, Hangzhou, Zhejiang Province, P.R. China
[2]School of Civil Engineering, Zhengzhou University, Zhengzhou, Henan Province, P.R. China

**Corresponding Author:**
Address correspondence to Yadan Yan: yanyadan@zzu.edu.cn

Compared with the large number of residential self-selection studies that have been conducted in developed countries, few studies have been performed in China. Because the high-density land use, special development process of the housing market, and specific stages of urbanization and motorization of China are significantly different from those of developed countries such as the U.S.A., conclusions from Western research may not be directly transferred to China (*2*).

This study aims to offer better understanding of residents' travel behavior especially the BEP in the Chinese context with an improved propensity score technique. It sets out to answer two questions: first, to what extent does the built environment exert causal influences on residents' travel behavior; second, what is the proportion of residential self-selection that brings about differences in observed residents' travel behavior. Based on Nanjing Household Travel Survey data, hierarchical clustering was used to classify residential built environments to multiple patterns and use propensity score weighting with generalized boosted models to control for residential self-selection. The observed effect (OBE), the average treatment effect on the population (ATE), and the BEP on travel behavior were then calculated.

The remainder of this paper is organized as follows. The second section reviews previous studies regarding the propensity score technique used in travel behavior studies. In the third section the data and variables used in this study are introduced. The fourth section provides a description of the methodology, including hierarchical clustering and propensity score weighting, followed by the model results, which are discussed in the fifth section. The final section summarizes the key findings and limitations.

## Literature Review

Assessing the influence of the built environment on travel behavior is a causal inference problem in statistics. Experimental studies are ideal for inferring causality, as random assignments of residents cancel out the effects of their socio-demographics and travel preferences on their travel behaviors. Conducting experimental studies on travel behaviors is unfeasible, however, for economical, ethical, and legal concerns. Researchers must reference causality through observational studies. Since residents choose where to live based on their preferences and constraints, residents in different built environments may also differ on socio-demographics and travel preferences which influence travel behavior (i.e., residential self-selection). For instance, residents who tend to travel by bus are more likely to self-select into neighborhoods where public transit is convenient; this tendency will affect their travel mode choices. Extensive studies have investigated the influence of the built environment on travel behavior, and a variety of approaches have been employed in these studies to control for residential self-selection.

Mokhtarian and Cao reviewed studies from the perspective of methodologies. Among the nine methodological solutions to residential self-selection, propensity score matching is highly recommended for cross-sectional data (*3*). Mokhtarian and Herick provided a detailed review focused on propensity score techniques, which is one of the methodologies most widely used to control for residential self-selection (*4*). When employing propensity score techniques, researchers built two models: a selection model, which is intended to estimate the effect of residential self-selection on the treatment variable, and an outcome model, which is intended to explore the effect of the treatment variable (and other potential covariates) on the outcome variable. The numerous propensity score methods in the literature differ in how they estimate the propensity scores and how they use the resulting estimated propensity scores.

Methodologically, the overwhelming majority of studies have relied on the parametric estimation of propensity scores via the multinomial logistic model, which assumes that covariates are linear and additive on the log-odds scale. The multinomial logistic model is less flexible and requires variable selection performed manually or through forward selection methods, which may omit important covariates, especially the polynomial or interaction terms, or inaccurately specify the functional forms, finally resulting in inaccurate propensity scores. Contrary to statistical approaches which assume a data model with parameters estimated from the data, machine learning tries to extract the relationship between an outcome and predictor through a learning algorithm without an a priori data model (*5*). Machine learning techniques have been suggested as promising alternatives to logistic regression for the estimation of propensity scores (*6*, *7*). Westreich et al. identified four techniques as alternatives to logistic regression: neural networks, support vector machines, decision trees, and meta-classifiers (in particular, boosting) and found meta-classifiers appear to be most promising for use in the context of propensity score analysis (*8*). Lee et al. examined the performance of various CART-based propensity score models using simulated data and found under conditions of both moderate non-additivity and moderate non-linearity, logistic regression had subpar performance, while boosted CART provided substantially best bias reduction (*9*). These results suggest that boosting or boosted CART is a better alternative to logistic regression. However, these methods have not been applied to estimate propensity scores in the travel behavior literature to our knowledge.

Most studies then use stratification or matching on the resulting estimated propensity score, as Rosenbaum and Rubin suggested (*10*), however, matching will discard any unmatched sample. Stratification is sensitive to the number of strata. Although a large number of strata may achieve reliable results, propensities within a given stratum may be too disparate for the assumption of equivalence to hold. Further, most studies involve just two groups of interest: one treatment and one control. The built environment is polarized to a binary treatment (usually urban versus suburban), and variability within the treatment is ignored. Several studies have extended the propensity score methods to multiple treatments with three or more conditions of interest. For M treatments, however, M*(M-1)/2 multinomial logistic models are required to estimate propensity scores. As M gets larger, matching or stratification will exponentially increase the burden on model estimation and balance assessment. The only exception the authors have noticed is the study of Parady et al., which extended the conventional binary treatment to a continuous urbanization-level treatment (*11*). More flexible approaches to modeling multinomial outcomes have received little attention.

As for the Chinese context, travel behavior studies which have explicitly addressed the residential self-selection issue remain considerably limited (*2*). Some studies in the Chinese context have produced confusing findings: better accessibility to public transit leads to more car use in non-work travel (*12*), and the land use characteristics of small street blocks, high-density street networks, and mixed-use are encouraging factors of automobile dependence (*13*). The effect of residential self-selection is ignored in these studies. Recently, a few studies have explored the effects of the built environment and residential self-selection on travel behavior in the Chinese context using different approaches. Cao and Yang developed a structural equations model to examine the effects of the built environment and residential self-selection on commuting trips and their related $CO_2$ emissions in Guangzhou (*14*). Wang and Lin employed a longitudinal study design linking the residential built environment to travel behavior and travel-related attitudes before and after home relocation in Beijing (*15*). Lin et al. adopted the natural experimental approach by dividing the Beijing samples into two sub-groups based on whether or not the respondents had much freedom when choosing where to live and examined the residential self-selection issue (*16*). Yang et al. applied propensity score technique to explore the effects of built environment on travel behavior between household members by controlling for residential self-selection (*17*). Chen et al. employed Heckman's sample selection model to examine the effectiveness of transit-oriented development on reducing personal vehicle kilometers traveled (VKT) (*18*). Zhang et al. adopted the natural experimental approach to study the effects of subway proximity on car ownership in Beijing (*19*). Among these six studies, only two quantified the relative contribution of the built environment and residential self-selection in influencing travel behavior. The last two studies were both focused on the influence of proximity to metro station. Even for the large cities in China, however, neighborhoods located around metro stations are very limited, hence studies focusing on neighborhoods without metro services are needed.

Nanjing is the capital of Jiangsu Province and the second largest city in the Yangtze River Delta, China. By focusing on the residents living in neighborhoods without metro services in Nanjing, this paper examines relationship between built environment and travel behavior and quantifies the relative contribution of the built environment itself and residential self-selection to travel behavior effects. Previous studies usually set two groups (one treatment and one control), built logistic regression models to predict propensity scores, and used matching or stratification to equalize propensity scores. This paper improves these conventional propensity score techniques by employing generalized boosted model (GBM) to predict propensity scores and weighting on the estimated propensity scores to balance covariates for multiple treatments.

## Data and Variables

### *Travel Behavior, Social-Demographic, and Travel Preference Variables*

This study used the 2012 Nanjing Household Travel Survey data, which were initially collected for the annual transportation statistics report. The data contained a total of 5,974 residents from 2,007 households. Given the purpose of the study, the respondents residing within a 1 km range of metro stations were excluded. In addition, the respondents who had logically inconsistent trips were excluded. After data cleaning, a total of 9,270 trip records for 3,698 residents from 1,332 households were screened out. The final data included household characteristics (e.g., car ownership, annual income, geographical coordinates of household location) for each household, personal characteristics (e.g., gender, age, occupation, education status, and driver's license), a one-day travel diary, and eight questions about the opinions on urban transportation for each household member.

Travel behavior was the dependent variable in this study, which chose three trip frequency variables (i.e., walking trip frequency, bicycle trip frequency, and public transit trip frequency) and one trip distance variable (VKT). These four variables are concerned with transportation or land use policy-making and are commonly used in the previous travel behavior studies cited for comparison.

Socio-demographics included gender, age, occupation, education status, driver's license, bus pass, household income, household car ownership, household bicycle ownership, the number of preschool children in the family, and the number of primary and secondary school students in the family. These variables were processed as follows: the age variable had eight choices and the household income variable had seven choices; they were converted to continuous variables by using the medians of each choice. The occupation variable had nine choices; it was merged into three choices (i.e., student, worker, or other) and taken as a nominal variable. The education status variable had four choices; it was merged into two choices (i.e., low education or high education) and taken as an ordinal variable. The gender, driver's license, and bus pass variables were treated directly as nominal variables. The household car ownership, household bicycle ownership, number of preschool children, and number of primary and secondary school students variables were directly treated as continuous variables.

The respondents were asked their opinions about urban transportation. Among the eight questions of opinions, two questions could reflect their travel preferences. The first question regarded their attitude towards the purchase and use of a private car: (a) encourage, (b) moderately restrict, (c) strictly restrict. The respondents' choices reflected their preferences on car use. These three choices were treated as an ordinal variable in the model. The second question regarded what they believed to be most important for improving urban transportation (up to two choices): (a) build more roads, (b) implement stricter management and punishment, (c) cultivate citizens' traffic awareness, (d) improve public transit, (e) build more parking lots, (f) improve walking environment. The respondents' choices reflected their preferences on travel mode choices. The last three choices were converted to three dummy variables (i.e., whether the individual prefers public transit, whether the individual prefers driving, and whether the individual prefers walking) in the model.

### Built Environment Variables

The classic indices describing residential built environment are the "5Ds" (i.e., density, diversity, design, destination accessibility, and distance to transit) proposed by Ewing and Cervero (*20*). This study used open data, such as Point of Interest (POI) and Walk Score, to measure the residential built environment by loosely following the "5Ds." Since the geographical coordinates for each household location were recorded, the residential built environments can be precisely measured.

POI represents a much finer-grained picture of land use than conventional land use data and have been shown to be a good proxy for community vitality (*21*). This study acquired approximately 26,000 POIs for Nanjing from Gaode—one of the largest web mapping and navigation providers in China. These POIs belonged to 64 subcategories of 12 primary categories. Three "D" indices were calculated using POIs. The density index included three types of POI densities, namely, the residence POI density, the work POI density, and the commerce POI density. The residence POIs and work POIs were two of the original primary categories. The commerce POIs were the sum of three primary original categories (i.e., the service POIs, the shop POIs, and the restaurant POIs). The three types of POI densities for each household were computed as the number of corresponding POIs within an 800 m buffer of household location. Since the numbers of the three types of POI densities were quite different, Equation 1 was used to obtain the normalized POI densities, which range from 0 to 1.

$$\mathrm{density_{ij}} = \frac{\mathrm{X}_{ij} - \min\left(\overrightarrow{\mathrm{X}_i}\right)}{\max\left(\overrightarrow{\mathrm{X}_i}\right) - \min\left(\overrightarrow{\mathrm{X}_i}\right)} \tag{1}$$

In the formula, $density_{ij}$ is the normalized type $i$ POI density of household $j$. $X_{ij}$ is the raw type $i$ POI density of household $j$. $\overrightarrow{\mathrm{Xi}}$ is the vector of the raw type $i$ POI densities of all households.

The diversity index, which describes the number of different land uses in a given area, was calculated using Equation 2 based on the three types of normalized POI densities (*22*).

$$\mathrm{diversity_j} = 1 - \frac{\frac{a-b}{d} - \frac{a-c}{d}}{2} \text{ with } \begin{cases} a = \max\left(\mathrm{density_{1j}}, \mathrm{density_{2j}}, \mathrm{density_{3j}}\right) \\ b = \min\left(\mathrm{density_{1j}}, \mathrm{density_{2j}}, \mathrm{density_{3j}}\right) \\ c = \mathrm{mean}\left(\mathrm{density_{1j}}, \mathrm{density_{2j}}, \mathrm{density_{3j}}\right) \\ d = \mathrm{sum}\left(\mathrm{density_{1j}}, \mathrm{density_{2j}}, \mathrm{density_{3j}}\right) \end{cases} \tag{2}$$

In the formula, $diversity_j$ is the diversity of household $j$, and $density_{1j,\ 2j,\ 3j}$ are the normalized residence, work, and commerce POI densities for household $j$, respectively.

The distance to transit index was measured by the density of bus stops within an 800 m buffer of household location. The bus stop POIs were one of the original subcategories. In the calculation, each bus stop was weighted by the number of bus routes serviced.

The design index always describes the street network characteristics within an area and the destination accessibility index always describes the ease of access to trip attractions. These two indices were measured using Walk Score. Walk Score is the most popular walkability rating system and has been widely used in the U.S.A., U.K., Canada, Australia, New Zealand, and other countries. The walk score of an address is calculated in two steps: first, the base score is computed based on the proximity to 13 amenity categories, considering distances, counts, and weights, which is closely related to the accessibility index; second, the base score may receive a penalty for having a poor walking environment such as long blocks or sparse intersections, which is closely related to the design index (*23*). Studies have shown that the walk score is consistent with the public understanding of walkability (*24*) and can be verified with the conventional walkability assessment method (*25*). Walk Score data can be obtained based on the geographical coordinates of the household location via the API provided by the walkscore.com website.

## Methodology

### *Hierarchical Clustering Analysis to Classify Residential Built Environments*

Because of the large geographical coverage and small sampling rate, the spatial distribution of respondents was sparse. If some typical neighborhoods were directly selected in which to investigate travel behavior, each neighborhood would not have enough respondents to cover different socio-demographics and travel preferences. As this study aimed to evaluate the impacts of the residential built environment holistically, it grouped the respondents with similar residential built environments, although they may be spatially dispersed.

The hierarchical clustering algorithm is one of the widely used cluster analysis methods. It does not need a cluster number as input but generates a dendrogram for users to determine the proper cluster number (*26*). Euclidean distance was used as the dissimilarity measure and the ward.D method as inter-cluster dissimilarity measure. This paper classified residential built environment based on the six variables by loosely following the "5Ds." Since the magnitudes of the six variables were quite different, all variables were standardized to make them obey the distribution with a mean of 0 and a standard deviation of 1. The situation in which large-scale variables have greater effects on clustering results than small-scale variables can thus be avoided. The standardization formula is shown in Equation 3.

$$\mathrm{X}' = \frac{\mathrm{X} - \mathrm{mean}(\vec{X})}{sd(\vec{X})} \tag{3}$$

In the formula, *X'* is the standardized variable. *X* is the raw variable before standardization, and $\vec{X}$ is the vector of the raw variable.

### *Propensity Score Weighting to Control for Residential Self-Selection*

As the residential self-selection effect exists, propensity score weighting was employed to correct for its sources. The propensity score is the conditional probability of assigning a participant to a treatment group rather than to a control group given multiple covariates which have combined effects. Conditional on the propensity score, covariates will be distributed equally in treatment and control groups and self-selection bias will be eliminated in non-randomized trials (*27*).

The application of propensity score weighting followed these steps: (a) estimate propensity scores, (b) assess balance, and (c) estimate weighted mean and causal treatment effects. Among these three steps, steps 2 and 3 are straightforward, while step 1 requires the greatest effort. Large numbers of covariates and uncertain functional forms prevent the accurate estimation of propensity scores. In this paper, a powerful machine learning method, GBM, was employed to obtain robust propensity score weights. GBM, first proposed by Friedman, is a general, automated, and data-adaptive nonparametric tree based technique (*28*). GBM will capture the flexible and nonlinear relationships between treatment assignment and the pretreatment covariates, does not need to exclude insignificant or collinear covariates, and can automatically handle covariate interaction terms and polynomial terms without over-fitting the data. While using multinomial logistic models, if the covariates are not successfully balanced on the first try, researchers need to add some polynomial or interaction terms of the unbalanced covariates into the model manually until all covariates are balanced. Improper parametric specification of relationships may lead to biased estimates. Another advantage is that GBM will automatically add indicators for missing values and include them in the model. Evidence also suggests that GBM outperforms the multinomial logistic model in relation to balance properties (*9*, *29*, *30*). The only potential drawback of GBM is that the model forms and parameters cannot be explicitly identified like many other machine learning methods. Since the model used to estimate propensity scores is built for predictive accuracy (in-sample fit, to be more specific) rather than interpretations of results, GBM is well suited. Friedman developed an extension of a variable's "relative influence" for GBM estimates (*28*), which means that GBM is not a fully "black box." For the mathematical details of GBM, readers can refer to a related paper (*31*).

As GBM uses a boosted algorithm to superimpose a series of weak learning classifiers to form a strong learning classifier, a stopping rule needs to be set to determine the optimal iteration for predicting the propensity scores. D'Agostino proposed using some form of means comparison on the covariates to test whether the covariates have been balanced (*32*). Similarly, in this study, the stopping rules for the optimal iteration were set when the maximum of the absolute standardized mean difference (ASMD) of the covariates between treatment groups is minimized. The final ASMD is also the balance metrics for each treatment. ASMD values of less than 0.20 are considered small, those of 0.40 are considered moderate, and those of 0.60 are considered large (*33*).

Propensity scores created by GBM are used to reweight cases from both treatment and control groups to match each other. Let $f(\mathbf{x}|t=1)$ be the distribution of features for the treatment cases and $f(\mathbf{x}|t=0)$ be the distribution of features for the control cases. As they differ from ATE, each group is weighted to match the population $f(\mathbf{x})$.

$$f(\mathbf{x}|t=1) = w(\mathbf{x})f(\mathbf{x}) \tag{4}$$

$$f(\mathbf{x}|t=0) = w(\mathbf{x})f(\mathbf{x}) \tag{5}$$

In the formula, $w(\mathbf{x})$ is the weight. Using Bayes Theorem we obtain $w(\mathbf{x}) \propto 1/f(t=1|\mathbf{x})$for the treatment group and $w(\mathbf{x}) \propto 1/f(t=0|\mathbf{x})$ for the control group. If $p(x_i)$ is the estimated propensity score for case $x_i$, the weight of treatment cases $w_i = 1/p(x_i)$ and the weight of control cases $w_i = 1/(1-p(x_i))$.

For multiple treatments, the GBM is used in the following fashion to obtain weights: first, create dummy indicators for each of the treatments; second, fit separate GBMs to each dummy indicator and obtain the propensity score; finally, compute the ATE weights, which are equal to the reciprocal of the probability that a respondent received the treatment, that is, the estimated propensity score from each of the GBM fits. Using the propensity score weighting to unify the sample composition of all groups, a virtual standard population is created. In this virtual standard population, the confounding factors among all groups tend to be similar, and these groups become comparable. If there are M treatments, only M separate GBMs are needed to obtain the weights. In contrast, propensity score matching or stratification requires M*(M-1)/2 multinomial logistic models (one model for each pair of control and treatment).

## Results

### *Hierarchical Clustering Results*

As shown in Figure 1, residents' household locations were clustered into four built environment patterns by hierarchical clustering. Although no variables involving geographical coordinates or spatial distances are used, the spatial distributions of the four built environment patterns have obvious spatial agglomeration characteristics. Pattern 1 is concentrated in the inner city. Pattern 2

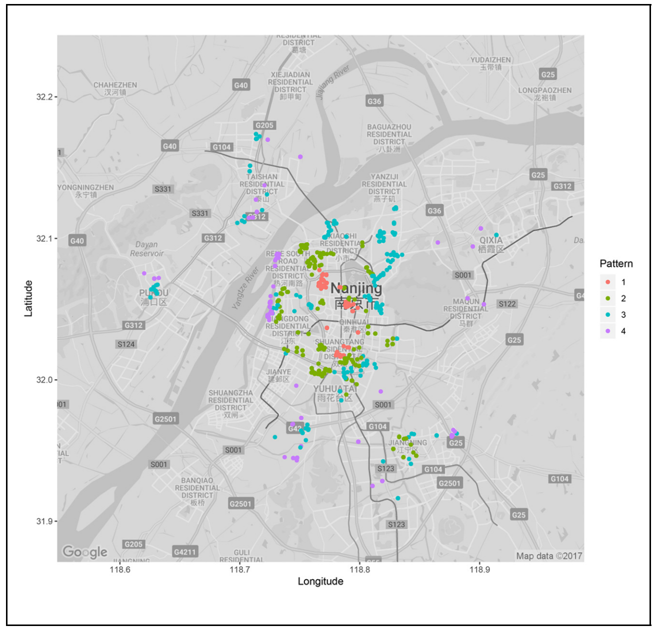


**Figure 1.** Four built environment patterns.

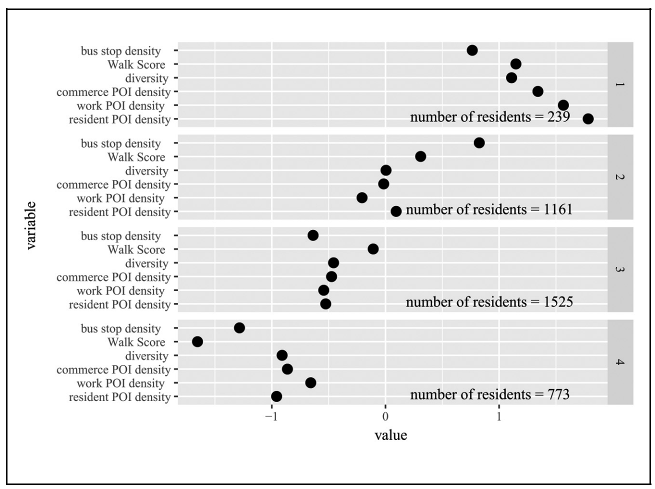


**Figure 2.** Average built environment characteristics of the four built environment patterns.

is located in the inner suburbs. Pattern 3 is distributed with some in the inner suburbs and some in the outer suburbs. Pattern 4 is located in the outer suburbs. The average distances to the city center increase in the order of Patterns 1, 2, 3, and 4 (2.60 km, 5.46 km, 9.76 km, and 13.16 km, respectively). Figure 2 is a Cleveland diagram showing the average built environment characteristics of the four patterns. The "5Ds" indices decrease in the same order. As the indices decline, the built environment changes from high-density and mixed land use with traditional street grid and high transit accessibility to low-density and single land use with large blocks, wide roads, and low transit accessibility.

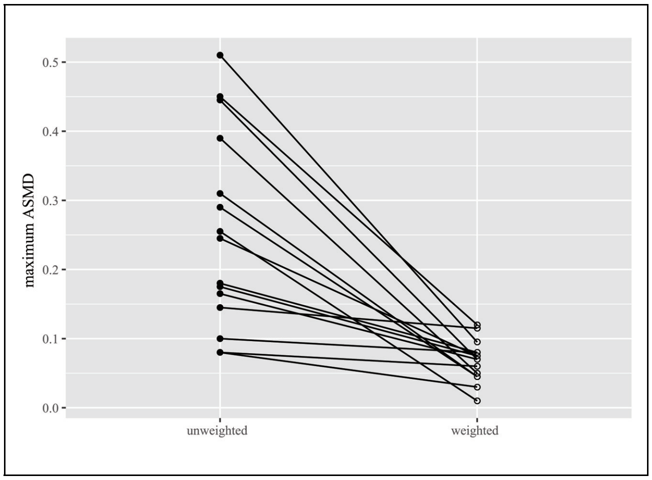


**Figure 3.** Relative influences of covariates on propensity scores.

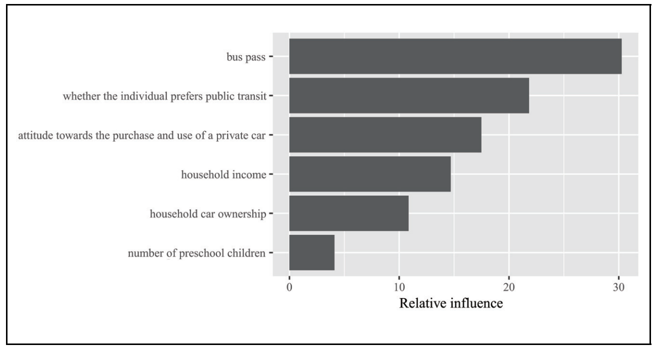


**Figure 4.** ASMD and t-tests of covariates before and after weighting.

## Propensity Score Weighting Results

The socio-demographics and travel preference variables were taken as pretreatment covariates since they were the confounding factors that affected both the residents' choices of travel behavior and household locations. Based on 15 socio-demographic and travel preference variables, four GBMs were built to predict propensity scores using R language.

The relative influences of the covariates can offer some insight into the contribution of covariates to the propensity scores. For tree based GBM the relative influence of a covariate $x_j$ is

$$\hat{J}_t^2 = \sum_{splits\ on\ x_j} I_t^2 \tag{6}$$

where $I_t^2$ is the empirical improvement by splitting on $x_j$ at that point. The relative influence of a covariate is to average the relative influence of covariate $x_j$ across all the trees generated by the boosting algorithm (*28*). The relative influences of the covariates (greater than 1) are shown in Figure 3. The greatest influences come from the bus pass, followed by whether the individual prefers public transit, their attitude towards the purchase and use of a private car, and their household income.

The ATE is the average expected difference in travel behaviors of the population residing in a particular built environment pattern versus that residing in another one. The population mean of the population residing in a particular built environment pattern equals the weighted mean of the residents who actually reside in the particular built environment pattern. To assess balance after applying the propensity score weights, ASMD and t-test were used for group differences in each covariate. One would expect the ASMD values to be less than 0.20 and the $p$-values of t-tests to be larger if balance had been achieved. Figure 4 shows the maximum ASMD of each covariate before and after weighting. The maximum values of ASMD decreased for all 15 covariates and are less than 0.2. The t-tests show that all covariates become insignificant (hollow circles in the figure) after weighting. After applying the propensity score weighting approach to balance the covariables, the distributions (or features of the distributions such as the means) of the covariables are similar across the treatment and control groups, thereby reducing the sample bias caused by residential self-selection.

## Effects of Residential Built Environment and Self-Selection

This study compared the walking trip frequency, bicycle trip frequency, public transit frequency, and VKT in the four built environment patterns, as shown in Table 1.

The mean travel behavior of the residents in each residential built environment pattern before and after weighting are shown in columns 3 and 4. The mean travel behavior of the population residing in a built environment pattern equals the mean travel behavior of the weighted residents who actually reside in that built environment pattern. Taking column 2 as the treatment conditions and columns 6, 7, and 8 as the control conditions, a lower triangular matrix (column 5) is created to record the OBE of built environment on travel behavior, the ATE of built environment on travel behavior, and the BEP. For each travel behavior outcome, six pairs of built environment patterns were obtained (i.e., 2-1, 3-1, 4-1, 3-2, 4-2, and 4-3). OBE is simply the "before" mean differences of the travel behavior outcome between the original treatment group and the control group; it was calculated as the average travel behavior outcome of the residents in the treatment built environment minus that of the residents in the control built environment. OBE is because of the combined effects of built environment and residential self-selection. ATE is the "after" mean differences of the travel

**Table 1.** OBE and ATE on Travel Behavior

| 1 Travel behavior indicators | 2 Built environment pattern | 3 Sample mean (unweighted) | 4 Sample mean (weighted) | 5 OBE / ATE / BEP | | |
|---|---|---|---|---|---|---|
| | | | | 6 Pattern 1 | 7 Pattern 2 | 8 Pattern 3 |
| Walking trip frequency | Pattern 1 | 0.983 | 0.854 | | | |
| | Pattern 2 | 0.629 | 0.634 | 0.354* / 0.220* / 62% | | |
| | Pattern 3 | 0.600 | 0.570 | 0.383* / 0.284* / 74% | 0.029 / 0.064 / — | |
| | Pattern 4 | 0.459 | 0.523 | 0.524* / 0.331* / 63% | 0.170* / 0.111* / 65% | 0.141* / 0.048 / 34% |
| Bicycle trip frequency | Pattern 1 | 0.946 | 1.021 | | | |
| | Pattern 2 | 0.924 | 0.959 | 0.022 / 0.062 / — | | |
| | Pattern 3 | 0.880 | 0.908 | 0.065 / 0.113 / — | 0.043 / 0.051 / — | |
| | Pattern 4 | 0.842 | 0.859 | 0.104 / 0.162 / — | 0.082 / 0.100 / — | 0.038 / 0.049 / — |
| Public transit trip frequency | Pattern 1 | 0.410 | 0.375 | | | |
| | Pattern 2 | 0.656 | 0.639 | –0.246* / –0.264* / 107% | | |
| | Pattern 3 | 0.641 | 0.608 | –0.231* / –0.233* / 101% | 0.015 / 0.031 / — | |
| | Pattern 4 | 0.525 | 0.521 | –0.115 / –0.146* / 127% | 0.131* / 0.118* / 90% | 0.116* / 0.087 / 75% |
| VKT (m) | Pattern 1 | 1014 | 1559 | | | |
| | Pattern 2 | 2189 | 2443 | –1174 / –884 / — | | |
| | Pattern 3 | 3093 | 3163 | –2079* / –1604* / 77% | –904* / –720 / 80% | |
| | Pattern 4 | 4509 | 3557 | –3495* / –1998* / 57% | –2321* / –1114* / 48% | –1416* / –395 / 28% |

*$p < 0.1$.

behavior outcome between the population in the treatment condition and control condition; it was calculated as the average travel behavior outcome of the weighted residents in the treatment built environment minus that of the weighted residents in the control built environment. ATE is only because of the effect of built environment. One-way ANOVA was used to identify overall significant differences between the residents of four built environment patterns, and TukeyHSD was used to further identify pairwise significant differences between each built environment pattern pair. The OBE or ATE marked with * in column 5 denote that the effect is significantly different from 0 at the 0.1 level. BEP is the proportion of OBE that can be attributable to ATE (BEP = ATE / OBE = built environment effect / (built environment effect + residential self-selection effect)) (*4*). The denominator constitutes the total effects of built environment and residential self-selection, and the numerator constitutes the true effect of built environment on travel behavior. Since the built environment and residential self-selection are expected to have the same sign, the BEP usually falls in the range of [0, 1]. However, if OBE and ATE are both insignificant in TukeyHSD, which means that the effects of built environment and residential self-selection on travel behavior are both near 0, then BEP is not estimated because of its meaninglessness. If either OBE or ATE is insignificant, then BEP is still estimated for reference, but it should be noted that BEP may have large random variation because one of the effects of built environment and residential self-selection is near 0.

For walking trip frequency, OBE and ATE are significant for most pairs of built environment patterns (five out of six and four out of six, respectively). Moving the population from a built environment pattern with lower "5Ds" indices to a built environment pattern with higher "5Ds" indices increases the walking trip frequency by 0.267 (40%) before weighting and by 0.176 (27%) after weighting on average. This confirms the finding from most previous studies that high-density, mixed-use, walkable, and transit-accessible neighborhoods make residents walk more. The BEPs range from 62% to 74% (we do not take BEP of the pair of Patterns 3 and 4 into consideration, since the insignificance of ATE makes the BEP unreliable). The BEPs are slightly larger than those in the research of Cao et al., who found that the built environment explained 61% of the observed influence of neighborhood type on the walking to store frequency (*34*). In general, walking trip frequency is much more a matter of built environment than residential self-selection.

For the bicycle trip frequency, OBE and ATE are both insignificant for all six pairs of built environment patterns, which means that the effects of either built environment or residential self-selection are negligible. Cycling is the most commonly used travel mode in China. The Nanjing Household Travel Survey shows the number of bicycle trips account for more than 33% of total trips. Further, approximately 60% of these bicycle trips use electric bicycles (e-bikes). The average travel distance of e-bikes is approximately 5 km, which is much longer than that of regular bicycles. A better residential built environment may induce residents to take regular bicycle trips, however, a worse residential built environment may compel residents to ride e-bikes to access services or take part in activities in more distant places. Zhao et al. found that among 12 built environment variables, only the number of public facilities within 300 m is positively associated with regular bicycle trips and only the number of bus stops with in 300 m radius is negatively associated with e-bike trips in Beijing (*35*), which confirms the argument of this study that the built environment may have opposite effects on regular bicycle trips and e-bike trips. When combining regular bicycle trips and e-bike trips together, Sun et al. found among 12 built environment variables that only one, the number of jobs per area in residential areas, was negatively associated with bicycle trips in Shanghai (*36*). These two studies also show very few built environment variables will affect bicycle trips, and this may be the other reason why the effect of built environment is negligible. A few studies obtain positive results by aggregating walking and cycling into active or non-motorized travel modes (*37*, *38*). Arguably this conventional practice is not suitable in the Chinese context because of the large share of bicycle trips.

For the public transit trip frequency, OBE and ATE are significant for about half of the pairs of built environment patterns (four out of six and three out of six, respectively). In general, residents living in neighborhoods with better built environment take more public transit trips. However, the residents living in Pattern 1 have the lowest public transit frequency even though Pattern 1 neighborhoods have the best built environment. We believe there are two opposite effects of built environment on the public transit frequency. Living in neighborhoods with good transit accessibility is an inducement to ride transit. Living in high-density, mixed-use, and walkable neighborhoods may provide short access distances to bus stops and link transit trips with errands on the way to and from bus stops, thus increasing the use of public transit, on the other hand, it may also lead to short travel distances, thus reducing the use of public transit. Since the effects of different built environment variables were not differentiated, these two effects are tangled. Only for the residents living in Pattern 1 is the latter effect dominant. However, if public transit trip frequency is combined with walking trip frequency and bicycle trip frequency, the total frequency of these green travel mode trips of the residents living in Pattern 1 is the highest among all the residents living in the four built environment patterns. A travel mode substitution mechanism from public transit to walking and cycling exists when the built environment gets better. Policies to shape the built environment may not necessarily increase public transit use but they can induce more sustainable travel behavior. Interestingly, the BEPs of moving the population from Patterns 4, 3, or 2 to Pattern 1 are slightly greater than 1, which means that residential self-selection is opposite in sign to built environment but smaller in magnitude than built environment. This result implies that residential mismatch exists. Although residents living in Pattern 1 prefer to use public transit, their travel preferences are suppressed by the built environment of their neighborhoods as their travel distances are always short. In general, a relatively large BEP (94%) was found, indicating that built environment plays a strongly dominant role in the impact on the public transit trip frequency (BEPs of the pair of Patterns 1 and 4 and the pair of Patterns 3 and 4 were not taken into consideration since the insignificance of OBE or ATE makes the BEP unreliable). This result appears to be inconsistent with previous studies. Lee et al. found that, for public transit commuting, only 57% of the observed influence can be attributable to built environment (*39*). Using public transit seems to be a necessity rather than a choice in the Chinese context.

For VKT, the effect of built environment is reduced after controlling for residential self-selection. On average, residents living in a better built environment pattern drove 1.9 km and 1.1 km less than those living in a worse built environment pattern before and after weighting. Among the six pairs of built environment patterns, five pairs are significant before weighting while only three pairs are still significant after weighting. Moving the population from a built environment pattern to its adjacent built environment pattern are all insignificant, which indicates reducing VKT requires quite large changes to the built environmental. BEPs range from 48% to 77%, results which are close to the findings in three previous articles. Mokhtarian and Herick summarized six studies quantifying the BEP of VKT and found that the BEPs ranged from 48% to 98% (*4*). Cao et al. identified five studies quantifying the BEP of VKT and found that the BEPs ranged from 38% to 76% (*40*). Chen et al. found 73% of the observed differences in VKT were attributed directly to the built environment itself in Shanghai (*18*).

## Conclusion

Using the 2012 Nanjing Household Travel Survey data, this paper explores the causal effects of built environment and residential self-selection on four travel behavior outcomes by hierarchical clustering and propensity score weighting. The presented methodology helps overcome some limitations of existing approaches in the travel behavior literature, particularly using GBM to replace simple multinomial logistic models to estimate propensity scores and using weighting to replace matching or stratification on the estimated propensity scores to balance multiple treatments.

Two main conclusions can be summarized from the study. First, this paper provides evidence that high-density, mixed-use, walkable, and transit-accessible built environment is associated with more walking trips and lower VKT but has no impact on bicycle trips and an inconsistent impact on public transit trips. As a holistic approach is adopted to characterizing residential built environments in relation to built environment patterns, the effects of some built environment variables are mixed, especially for bicycle and public transit trips. Modifying the residential built environment would not necessary increase bicycle trips or public transit trips but would indeed lead to meaningful and positive changes in travel behavior, for example, lower VKT and greater use of alternative travel modes. Again, reducing VKT requires quite large changes to the built environmental. Second, most travel behavior differences are primarily a result of built environment, but considering residential self-selection is still quite important. BEPs range from 62% to 74% for walking trip frequency, from 90% to 107% for public transit frequency, and from 48% to 77% for VKT (only considering the pairs of patterns where OBE and ATE are both significant). The average BEPs of VKT are less than those of walking or public transit trip frequency, and the variation in the BEPs of VKT is larger than those of walking or public transit trip frequency. Thus, among the four travel behavior outcomes, residential self-selection has the most influence on VKT. Because of data limitation, this paper does not consider the housing types and familial attitudes of residential choice when accounting for the effect of residential self-selection. Diversified housing types have different implications for residential choice, which will make the interactions between residential self-selection and travel behavior more complex. Based on the methodology framework proposed in this paper, it would not be difficult to extend this study further by adopting more activity-based variables or considering other aspects of travel behavior, which would be important future works.

## Acknowledgments

This study was funded by the National Nature Science Foundation of China (51608483) and the Zhejiang Philosophy and Social Science Planning Project (19NDQN351YB).

## Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: YD; data collection: YY; analysis and interpretation of results: YD; draft manuscript preparation: YD, YY. Both authors reviewed the results and approved the final version of the manuscript.

## References

1. Cao, X. J., P. L. Mokhtarian, and S. L. Handy. Examining the Impacts of Residential Self-selection on Travel Behaviour: a Focus on Empirical Findings. *Transport Reviews*, Vol. 29, No. 3, 2009, pp. 359–395.
2. Wang, D., and M. Zhou. The Built Environment and Travel Behavior in Urban China: a Literature Review. *Transportation Research Part D: Transport and Environment*, Vol. 52, 2017, pp. 574–585.
3. Mokhtarian, P. L., and X. Cao. Examining the Impacts of Residential Self-selection on Travel Behavior: a Focus on Methodologies. *Transportation Research Part B: Methodological*, Vol. 42, No. 3, 2008, pp. 204–228.
4. Mokhtarian, P. L., and D. Van Herick. Quantifying Residential Self-selection Effects: a Review of Methods and Findings from Applications of Propensity Score and Sample Selection Approaches. *Journal of Transport and Land Use*, Vol. 9, No. 1, 2016, pp. 9–28.
5. Breiman, L. Statistical Modeling: the Two Cultures (with Comments and a Rejoinder by the Author). *Statistical Science*, Vol. 16, 2001, pp. 199–215.
6. Setoguchi, S., S. Schneeweiss, M. A. Brookhart, R. J. Glynn, and E. F. Cook. Evaluating Uses of Data Mining Techniques in Propensity Score Estimation: A Simulation Study. *Pharmacoepidemiology and Drug Safety*, Vol. 17, No. 6, 2008, pp. 546–555.
7. McCaffrey, D. F., G. Ridgeway, and A. R. Morral. Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies. *Psychological Methods*, Vol. 9, No. 4, 2004, pp. 403–425.
8. Westreich, D., J. Lessler, and M. J. Funk. Propensity Score Estimation: Neural Networks, Support Vector Machines, Decision Trees (Cart), and Meta-Classifiers as Alternatives to Logistic Regression. *Journal of Clinical Epidemiology*, Vol. 63, No. 8, 2010, pp. 826–833.
9. Lee, B. K., J. Lessler, and E. A. Stuart. Improving Propensity Score Weighting Using Machine Learning. *Statistics in Medicine*, Vol. 29, No. 3, 2009, pp. 337–346.
10. Rosenbaum, P. R., and D. B. Rubin. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, Vol. 79, No. 387, 1984, pp. 516–524.
11. Parady, G. T., K. Takami, and N. Harata. Connection between Built Environment and Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2453: 137–144.
12. Ta, N., Y. Chai, and M. Guan. The Relationship between the Built Environment and Car Travel Distance on Weekdays in Beijing. *Acta Geographica Sinica*, Vol. 70, No. 10, 2015, pp. 1675–1685.
13. Wei, Y., and C. Pan. Urban Land-use Characteristics and Commuters' Travel Pattern: A Case Study of West Hangzhou. *City Planning Review*, Vol. 36, No. 3, 2012, pp. 76–84.
14. Cao, X., and W. Yang. Examining the Effects of the Built Environment and Residential Self-selection on Commuting Trips and the Related $CO_2$ Emissions: an Empirical Study in Guangzhou, China. *Transportation Research Part D: Transport and Environment*, Vol. 52, 2017, pp. 480–494.

15. Wang, D., and T. Lin. Built Environment, Travel Behavior, and Residential Self-selection: a Study Based on Panel Data from Beijing, China. *Transportation*, Vol. 43, 2017, pp. 1–24.
16. Lin, T., D. Wang, and X. Guan. The Built Environment, Travel Attitude, and Travel Behavior: Residential Self-selection or Residential Determination? *Journal of Transport Geography*, Vol. 65, 2017, pp. 111–122.
17. Huang, X., X. J. Cao, X. Cao, and J. Yin. How Does the Propensity of Living near Rail Transit Moderate the Influence of Rail Transit on Transit Trip Frequency in Xi'an? *Journal of Transport Geography*, Vol. 54, 2016, pp. 194–204.
18. Chen, F., J. Wu, X. Chen, and J. Wang. Vehicle Kilometers Traveled Reduction Impacts of Transit-oriented Development: Evidence from Shanghai City. *Transportation Research Part D: Transport and Environment*, Vol. 55, 2017, pp. 227–245.
19. Zhang, Y., S. Zheng, C. Sun, and R. Wang. Does Subway Proximity Discourage Automobility? Evidence from Beijing. *Transportation Research Part D: Transport and Environment*, Vol. 52, 2017, pp. 506–517.
20. Ewing, R., and R. Cervero. Travel and the Built Environment: A Meta-analysis. *Journal of the American Planning Association*, Vol. 76, No. 3, 2010, pp. 265–294.
21. Yue, Y., Y. Zhuang, A. G. O. Yeh, J. Xie, C. Ma, and Q. Li. Measurements of Poi-based Mixed Use and their Relationships with Neighbourhood Vibrancy. *International Journal of Geographical Information Science*, Vol. 31, No. 4, 2016, pp. 1–18.
22. Reusser, D. E., P. Loukopoulos, M. Stauffacher, and R. W. Scholz. Classifying Railway Stations for Sustainable Transitions - Balancing Node and Place Functions. *Journal of Transport Geography*, Vol. 16, No. 3, 2008, pp. 191–202.
23. *Walk Score Methodology*. Walk Score®, 2011.
24. Carr, L. J., S. I. Dunsiger, and B. H. Marcus. Walk Score™ as a Global Estimate of Neighborhood Walkability. *American Journal of Preventive Medicine*, Vol. 39, No. 5, 2010, pp. 460–463.
25. Carr, L. J., S. I. Dunsiger, and B. H. Marcus. Validation of Walk Score for Estimating Access to Walkable Amenities. *British Journal of Sports Medicine*, Vol. 45, No. 14, 2011, pp. 1144–1148.
26. Hastie, T., R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The Elements of Statistical Learning*, 2nd ed. Springer, Berlin, 2016.
27. Rosenbaum, P. R., and D. B. Rubin. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, Vol. 70, No. 1, 1983, pp. 41–55.
28. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, Vol. 29, No. 5, 2001, pp. 1189–1232.
29. Lee, B. K., J. Lessler, and E. A. Stuart. Weight Trimming and Propensity Score Weighting. *PLoS One*, Vol. 6, No. 3, 2011, p. e18174.
30. Harder, V. S., E. A. Stuart, and J. C. Anthony. Propensity Score Techniques and the Assessment of Measured Covariate Balance to Test Causal Associations in Psychological Research. *Psychological Methods*, Vol. 15, No. 3, 2010, pp. 234–249.
31. Elith, J., J. R. Leathwick, and T. Hastie. A Working Guide to Boosted Regression Trees. *Journal of Animal Ecology*, Vol. 77, No. 4, 2008, pp. 802–813.
32. D'Agostino, R. J. Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group. *Statistics in Medicine*, Vol. 17, No.19, 1998, pp. 2265-2281.
33. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Routledge, New York, 1988.
34. Cao, X. J. Exploring Causal Effects of Neighborhood Type on Walking Behavior Using Stratification on the Propensity Score. *Environment and Planning A*, Vol. 42, No. 2, 2010, pp. 487–504.
35. Zhao, C., T. A. S. Nielsen, A. S. Olafsson, T. A. Carstensen, and X. Meng. Urban Form, Demographic and Socio-Economic Correlates of Walking, Cycling, and e-Biking: Evidence from Eight Neighborhoods in Beijing. *Transport Policy*, Vol. 64, 2018, pp. 102–112.
36. Sun, B., A. Ermagun, and B. Dan. Built Environmental Impacts on Commuting Mode Choice and Distance: Evidence from Shanghai. *Transportation Research Part D: Transport and Environment*, Vol. 52, 2017, pp. 441–453.
37. Khan, M., K. M. Kockelman, and X. Xiong. Models for Anticipating Non-Motorized Travel Choices, and the Role of the Built Environment. *Transport Policy*, Vol. 35, 2014, pp. 117–126.
38. Yang, S., Y. Fan, W. Deng, and L. Cheng. Do Built Environment Effects on Travel Behavior Differ between Household Members? A Case Study of Nanjing, China. *Transport Policy*, Vol. 60, 2017, pp. 1–17.
39. Lee, J. S., P. Christopher Zegras, E. Ben-Joseph, and S. Park. Does Urban Living Influence Baby Boomers' Travel Behavior? *Journal of Transport Geography*, Vol. 35, 2014, pp. 21–29.
40. Cao, X. J., Z. Xu, and Y. Fan. Exploring the Connections among Residential Location, Self-selection, and Driving: Propensity Score Matching with Multiple Treatments. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 10, 2010, pp. 797–805.

*The Standing Committee on Traveler Behavior and Values (ADB10) peer-reviewed this paper (19-01002).*