The Journal of Transport and Land Use http://jtlu.org
Vol. 10 No. 1 [2017] pp. 763–788

Built environment and travel behavior: Validation and application of a continuous-treatment propensity score stratification method

Giancarlos Troncoso Parady
University of Tokyo
gtroncoso@ut.t.u-tokyo.ac.jp

Kiyoshi Takami
University of Tokyo
takami@ut.t.u-tokyo.ac.jp

Noboru Harata
University of Tokyo
nhara@ut.t.u-tokyo.ac.jp

Abstract: This article discusses the validation and implementation of a propensity score approach with continuous treatment to test the existence of a causal relationship between the built environment and travel behavior using cross-sectional data. The implemented methodology differs from previous applications in the planning literature in that it relaxes the binary treatment assumption, which polarizes the built environment into two extremes (e.g., urban vs suburban). The effectiveness of the proposed methodology in reducing bias was validated via Monte Carlo simulation. The proposed approach was shown to reduce self-selection bias against Ordinary Least Squares (OLS) regression in all but extreme levels of non-linearity. Empirical results suggest that an increase in urbanization has a negative effect on home-based maintenance car trip frequencies, and conversely, a positive effect on home-based maintenance non-motorized trip frequencies. Result estimates suggest the existence of a causal mode substitution mechanism between car and non-motorized modes given increases in the urbanization level at residential locations, thus providing some empirical support to the arguments put forth by compact city advocates.

Article history:
Received: February 7, 2016
Received in revised form: December 24, 2016
Accepted: March 3, 2017
Available online: August 24, 2017
Data availability: http://www.ut.t.u-tokyo.ac.jp/repository/gian/JTLU2017_data.zip

1 Introduction

Against the backdrop of urban sprawl and suburbanization, worsening traffic conditions and declining city centers, recent years have seen a paradigm shift in the conceptualization of what constitutes good urban development. Be it New Urbanism Smart Growth, or Compact Cities, one of the main premises behind these new paradigms is that mixed-use, high density developments can reduce automobile dependency and promote alternative modes such as transit, bicycles or walking.

The underlying assumption behind this premise is that there exists a non-spurious, causal mechanism behind the built environment and travel behavior connection. Therefore, the main objective of

http://dx.doi.org/10.5198/jtlu.2017.890
ISSN: 1938-7849

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 article is to test the existence or not of this causal mechanism, focusing in particular on tackling residential self-selection bias. More specifically, this study seeks to answer the following research questions:

- Does the built environment, as measured by urbanization level at one's residential location, has a causal effect on maintenance trip frequencies by mode? If so, what is the nature of this effect?
- For maintenance trips, does a mode substitution effect exist between car and non-motorized modes given changes in the urbanization level at one's residential location?

In particular, given the scarce nature of panel data, this study focuses on establishing causality using more widely available cross-sectional data. To do so, a propensity score approach is implemented using a continuous treatment variable as proposed by Troncoso Parady, Takami, and Harata (2014a). The proposed approach consists of a three-step process: (i) the estimation of a continuous urbanization level index (i.e., treatment of interest). (ii) The estimation of a continuous propensity score using the estimated urbanization index as treatment variable and (iii) the estimation of urbanization level average treatment effects on travel behavior through stratification on the propensity score. By using a continuous treatment variable, this approach does without the arbitrary classification of the built environment into "urban" or "suburban," thus accounting for the variability in the urbanization level of cities. The analysis is conducted using data from an original survey conducted in Fukuoka City, Japan. The performance of the continuous treatment propensity score method in terms of residential self-selection bias-reduction is validated through Monte Carlo simulation.

The rest of the paper is structured as follows. Section 2 provides an overview of existing findings in the residential self-selection literature. Section 3 elaborates on the methodological aspects of this article, including an overview of the propensity score approach (3.1), the generalization to continuous treatments (3.2), methodological comparison through Monte Carlo simulation (3.3), and the continuous treatment estimation (3.4). Section 4 details the general characteristics of the data used to test the study hypotheses, while Section 5 summarizes the modeling results. Finally, Section 6 wraps up the main conclusions of the article, its policy implications and limitations.

2 Literature review

The residential self-selection problem stems from households choosing their residential location, partly as a result of travel-related preferences and attitudes. If not controlled for, the effect of these preferences and attitudes might be confounded with the built environment effect, resulting in biased estimates of the true built environment effect. A considerable number of studies have addressed the residential self-selection problem. Since the literature has been widely documented elsewhere (Cao, Mokhtarian, & Handy, 2009a; Troncoso Parady, Chikaraishi, Takami, Ohmori, & Harata, 2015) only a brief outlook is provided here, specifically focusing on studies analyzing trip/tour frequencies, unless otherwise stated. The reader is referred to Bohte, Matt, and van Wee (2009) and Scheiner (2014) for a comprehensive discussion on the role of attitudes and preferences on residential self-selection, and to Naess (2014) for a counterargument on the relevance of the problem to the research community.

From a cross-sectional approach, one source of self-selection bias is variable omission. In that sense, bias can be mitigated by including in the deterministic component of the model equation the variables associated with residential location, such as preferences and attitudes, as well as other socio-demographics. This approach is referred to by Mokhtarian and Cao (2008) as the statistical control approach. After accounting for attitudes and preferences, Kitamura, Mokhtarian, and Laidet (1997) found that these factors explained a higher proportion of observed trip frequencies, and controlling for them reduced the magnitude of the land use effect. It is important to note; however, that attitudes and preferences do not render the built environment effect insignificant (Chatman, 2009). Using a similar strategy, strong effects have been observed particularly for non-motorized (NMM) trips, suggesting the existence of a mode substitution mechanism with private vehicles (Cao, Handy, & Mokhtarian, 2006; Cao et al., 2009b; Naess, 2009). The statistical control approach; however, is limited by the uncertainty of the effectiveness of the covariates used, especially in the case of attitudes, where there is no overarching theory guiding the definition and measurement of attitudes. This results in a wide variety of attitudinal measures that differ from study to study and make it difficult to compare results and evaluate accuracy of findings (Bohte et al., 2009).

Khattak and Rodriguez (2005) found via an instrumental variable approach, that households in neo-traditional neighborhoods exhibit less car trips and shorter distances, even though overall trip frequencies are similar. Boarnet and Sarmiento (1998) used (i) the percentage of buildings built before 1940 and, (ii) the percentage of buildings built before 1960 as instruments for the built environment, given that housing stock characteristics are likely correlated with land use patterns, yet uncorrelated with transport. They found no significant effects in most models and high sensitivity to model specification. On the other hand, using similar instruments, Vance and Hedel (2007) found evidence backing the existence of a casual mechanism between urban form and car use, and robustness to alternative model specifications. In spite of all, finding a proper instrument can be a difficult task.

From a quasi-longitudinal approach, changes in perception of accessibility have been associated with driving and walking level changes (Handy, Cao, & Mokhtarian, 2005; 2006). SEM studies have also found evidence of mode substitution with higher level of car use and lower levels of transit use associated with suburban relocation (Scheiner & Holz-Rau, 2013), and reduced driving associated with relocation to neo-traditional neighborhoods (Cao, Mokhtarian, & Handy, 2007). The main limitations of this approach, however, are the risk of recall error and the difficulties associated with measuring attitudes in the past (Cao et al., 2007).

Finally, from a longitudinal approach, using first-differenced OLS regressions Krizek (2003) found that as neighborhood accessibility increases, number of household tours increase, yet driven distances decrease. Troncoso Parady, Chikaraishi, Takami, and Harata (2014b) found via a fixed effect model, evidence of substitution effect between nearby activities reached by non-motorized modes and faraway activities reached by car, given accessibility level changes at home location. Although ideal due to its proximity to an experimental situation, true panel data studies in the literature are rather few in number due mostly to data collection difficulties.

2.1 Propensity score applications in the planning literature

Although not extensively, several studies in the transport literature have implemented propensity score methodologies as a way to address the residential self-selection problem. In a non-randomized treatment assignment context, its attractiveness derives from the potential to remove bias stemming from a large set of observed covariates $\mathbf{X}_i$ using a single scalar function (Rosenbaum & Rubin, 1983). In the case of residential self-selection this is done by (i) estimating the residential location choice probability and (ii) using the estimated probabilities to estimate built environment effects on travel behavior via propensity score matching, weighing or stratification.

Empirical findings suggest that even after controlling for residential self-selection, positive relations exist between vehicle kilometers driven and distance from the city center (Cao, Yu, & Fan, 2010), and between higher levels of business diversity and four-way intersections with more walking (Boer, Zheng, Overton, Ridgeway, & Cohen, 2007). In addition individuals living in neo-traditional neighborhoods were found to walk more than those living in suburban areas (Cao, 2010).

These studies highlight the potential of the propensity score approach to mitigate selection bias. However, to the best of our knowledge, all propensity score studies addressing the residential self-selection problem polarized the built environment to a binary treatment (usually urban vs. suburban), ignoring the inherent variability in terms of how "urban" or how "suburban" a neighborhood is. In that sense, the continuous approach discussed in this article allows for the estimation of the average treatment effect by taking into consideration the full spectrum of variability in the urbanization level across a city, doing without the need to arbitrarily define what "suburban" or "urban" means.

3 Methodology

3.1 Propensity score function and treatment estimators: The binary treatment case

Rosenbaum and Rubin (1983) defined the propensity score function as the conditional probability of treatment given observed covariates. The theoretical basis supporting the propensity score are discussed in detail in Rosenbaum and Rubin, but are briefly summarized here in order to provide a general understanding of the concept at hand.

- *The propensity score as a balancing score:* Given a binary treatment z, as a function of observed covariates the propensity score will balance $\mathbf{X_i}$, so that conditional on the propensity score function $P(\mathbf{X_i})=Pr(z_i|\mathbf{X_i})$, the distribution of $\mathbf{X_i}$ is the same for treated and untreated groups. In other words, conditional on $P(\mathbf{X_i})$, $\mathbf{X_i}$ and z are independent

$$Pr\{z_i|\mathbf{X_i}, P(\mathbf{X_i})\}=Pr\{z_i|P(\mathbf{X_i})\} \quad (1)$$

- *The strong ignorability assumption:* Given equation (1), strong ignorability of treatment implies that outcomes (Y_{0i}, Y_{1i}) are independent from treatment assignment given $P(\mathbf{X_i})$. In addition, every unit has a chance to receive either treatment state

$$Pr\{(Y_{0i}, Y_{1i})|z_i, P(\mathbf{X_i})\}=Pr\{(Y_{0i}, Y_{1i})|P(\mathbf{X_i})\};\ 0<Pr(z_i=1|P(\mathbf{X_i})<1 \quad (2)$$

Rosenbaum and Rubin (1983) note that in a randomized trial the propensity score is a known function defined by the randomization mechanism. In a nonrandomized case; however, this function is not known but can be estimated from observed data, using limited dependent variable models such as the logit model in the case of discrete choices. Care should be taken to include as many relevant covariates as possible in the specification function.

Given that the two conditions above hold, Rosenbaum and Rubin (1983) show that at any value of the balancing score, the difference between the treatment and control means is an unbiased estimate of the average treatment effect at the value of the balancing score. As such, unbiased estimates of treatment effects can be estimated via several estimators. To do so, several approaches have been proposed, of which the most common are matching (Heckman, Ichimura, Smith, & Todd, 1998), weighting (Horvitz & Thompson, 1952; Imbens & Wooldridge, 2008), and stratification (Rosenbaum & Rubin, 1984). Of these three, stratification is the most relevant method to this study, as it can be easily adapted to the continuous treatment case.

The stratification approach consists on sub-classifying the sample on J number of strata based on the propensity score where the ATE can be estimated as

$$ATE_{stratification}=\Sigma_{j=1}^{J}(\overline{Y}_{j1}-\overline{Y}_{j0})\cdot W_j \quad (3)$$

where $\overline{Y}_{j1}$ is the mean outcome in class j when treated, $\overline{Y}_{j0}$ the mean outcome in class j when untreated, and W_j is the relative weight of strata j estimated as n_j/N. Rosenbaum and Rubin(1984) showed that a 5 strata sub-classification of the propensity score might reduce over 90% of bias due to observed covariates. Imbens and Wooldridge (2008) point out; however, that although five strata have been commonly used empirically, depending on sample size and the joint distribution of the data, fewer or more strata might results in lower mean square error.

3.2 Generalizing the propensity score to continuous treatments

A generalization of the propensity score method was proposed by Imai and van Dyk (2004) to allow for arbitrary treatment regimes T_i^A. Following Imai and van Dyk, the distribution of a continuous treatment T_i^A given a vector of covariates $\mathbf{X_i}$, is modeled as $T_i^A \mid \mathbf{X_i} \sim N(\mathbf{X_i}^{\top}\beta,\sigma^2)$. The propensity score function $P(\mathbf{X_i})=\Pr\{T_i^A \mid \theta_\psi(\mathbf{X_i})\}$ is assumed Gaussian distributed, and parameterized by $\psi=(\beta,\sigma^2)$, so that $\theta_\psi(\mathbf{X_i})=\mathbf{X_i}^{\top}\beta$. This implies that the propensity score function is solely characterized by the scalar θ, and its estimator $\hat{\theta}_\psi(\mathbf{X_i})=\mathbf{X_i}^{\top}\hat{\beta}$, is uniquely characterized by the conditional mean function of the linear regression of the treatment variable $T_i^A=t^P$ and all covariates $\mathbf{X_i}$, where t^P is a potential treatment.

It can also be shown that for non-binary treatments, the propensity score is as a balancing score

$$\Pr\{T_i^A \mid \mathbf{X_i}, P(\mathbf{X_i})\}=\Pr\{T_i^A \mid P(\mathbf{X_i})\} \tag{4}$$

and that given $P(\mathbf{X_i})$ the outcome distribution of a potential treatment t^P, $Y_i(t^P)$ is independent from treatment assignment

$$\Pr\{Y_i(t^P) \mid T_i^A, P(\mathbf{X_i})\}=\Pr\{Y_i(t^P) \mid P(\mathbf{X_i})\} \tag{5}$$

for any $t^P \in T$, where T is a set of potential treatment values. Thus, by averaging $\Pr\{Y_i(t^P) \mid P(X_i)\}$ over the distribution of $P(\mathbf{X_i})$, the distribution of the outcome of interest can be obtained as

$$\Pr\{Y_i(t^P)\}=\int \Pr\{Y_i(t^P) \mid T_i^A=t^P,\theta\} \Pr(\theta)d\theta. \tag{6}$$

This integration can then be approximated parametrically as $\Pr_\phi\{Y_i(t^P) \mid T_i^A=t^P\}$ stratified by the propensity score θ, where ϕ parameterizes the distribution. Thus, the distribution of $Y_i(t^P)$ can be approximated as the weighted average of the within strata outcome distribution

$$\Pr\{Y_i(t^P)\}\approx\Sigma_{j=1}^{J} \Pr_{\hat{\phi}_j}\{Y_i(t^P) \mid T_i^A=t^P\}\cdot W_j \tag{7}$$

where $\hat{\phi}_j$ is the within strata estimate of unknown parameter ϕ in strata *j*, and W_j is the relative weight of strata *j*. ϕ can then be estimated as

$$\hat{\phi}=\Sigma_{j=1}^{J}\hat{\phi}_j\{Y_i(t^P) \mid T_i^A=t^P,\mathbf{X_i}\}\cdot W_j \tag{8}$$

where covariates $\mathbf{X}_i$ are included to control for variability of θ within strata. The average treatment effect is then a function of $\hat{\phi}$; in this case, the weighted treatment coefficient of the regression of the outcome variable $Y_i(t^P)$ on t^P and all covariates, where weights are given by the sample relative weight n_j/N. Variance for the weighted coefficients can be estimated as

$$\Sigma_{j=1}^{J} W_j^2 \cdot Var(\hat{\beta}_j) \tag{9}$$

where W_j is the weight of each strata j, and $\sum_{j=1}^{J} W_j=1$.

3.3 Methodological comparison through simulation

The performance of the propensity score methodology is tested against the OLS full-covariate model (statistical control approach) through Monte Carlo simulation. Two set of simulations are estimated, corresponding to home-based maintenance trips by car and by non-motorized means. Although relevant covariates related to travel behavior and residential location are known to some extent, the true data generating process is unknown, in that sense, Following Rubin & Thomas (2000) and Imai and van Dyk (2004), an exponential function was used to specify a multiplicative data generating processes (DGP) with different levels of linearity. Departing from Imai and van Dyk, the data generating process is of the form

$$\mathrm{Y_i}=\delta_i\, T_i^A+c(\lambda)e^{\Sigma_{k=1}^{K}\boldsymbol{\lambda_k X_{ik}}} \qquad (10)$$

where for the *i*th individual, $\mathrm{Y_i}$ is the simulated outcome (e.g., home-based maintenance trip frequencies by mode), δ_i is the treatment effect, T_i^A is the assigned treatment, and λ_k is a vector of zero-mean Gaussian distributed coefficients for a vector of covariates $\mathbf{X_i}$ of *k* dimensions. The variance of λ_k is then used to control the level of linearity of each model. Each simulation was run with 1000 replications. The constant $c(\lambda)$ is defined such that the variable of the simulated outcome variable approaches the variable of the observed outcome variable.

The degree of linearity of each simulation is measured by the average R^2 of the simulated outcomes regressed on the treatment variable and the set of covariates[1] . Four levels of linearity are considered: (i) extreme non-linearity with R^2 values ranging from 0.10 to 0.15, (ii) high non-linearity with R^2 values ranging from 0.25 to 0.30, (iii) non-linearity with R^2 values ranging from 0.40 to 0.45 and (iv) moderate non-linearity with R^2 values ranging from 0.65 to 0.70. In travel demand models R^2 values usually range between levels (i) and (ii) and less often near level (iii) values.

As in Rosenbaum and Rubin (1984) and Imai and van Dyk (2004), the simulations are conducted under the assumption that the true propensity score function in known.

3.4 Defining the treatment of interest: A continuous index of urbanization

Urbanization level at the location of residence, measured as a continuous variable, was defined as the treatment variable of interest. In order to quantify urbanization level, a latent variable model was specified using confirmatory factor analysis (CFA). CFA allows for a complete specification of the nature of relation between the latent factor and its indicators, as well as for the calculation of goodness of fit statistics (Brown, 2006).

3.4.1 The spatial analysis unit

A critical part of the analysis is the definition of the basic spatial unit. Particularly due to the modifiable areal unit problem (MAUP), a pervasive yet widely ignored problem in spatial analysis, stemming from the way spatial data is aggregated. This problem, as argued by Fotheringham and Wong (1991) might have unpredictable effects in multivariate analysis. Given that spatial zones in widely used datasets such as the national census are defined rather arbitrarily, how sensitive are estimated results to changes in terms of zoning and scale is a non-trivial problem. Empirical research; however, has shown that a regular aggregation scheme such as a rectangular tessellation tends to produce more tractable results than aggregation on census geographic units (Putman & Chung, 1989; Zhang & Kukadia, 2005). Accordingly, to address the zonal problem, instead of the existing political district divisions, a regular sampling scheme is

[1] Covariates are fixed among all replications as the observed values in the dataset are used.

implemented. A 300m wide hexagon (150m from the center to any vertex) tessellation was used to subdivide the city area in regular spatial units. Although more common in ecological modelling, a hexagonal grid was selected as it presents some advantages over the rectangular grid, such as a better match in Euclidian distance measurements, and greater clarity in visualization (Birch, Oom, & Beecham, 2007).

Regarding the aggregation scale problem, as suggested by Jelisnki and Wu (1996) and Dark and Bram (2007) a sensitivity analysis was conducted in order to analyze how sensitive results are to variations in the scale of analysis. Therefore, in addition to the 300m wide hexagon, three additional scales were used for the sensitivity analysis; 100m, 600m and 1000m wide hexagons (Sensitivity analysis results not included here, but are available upon request to the authors).

3.4.2 Definition of the indicator variables

In urban economics, combination of factors such as resource and transport advantage, economies of scale, and preference for variety in consumption and production are commonly agreed to give way to the urban agglomerations (Fujita, 1989). A myriad of factors such as land use allocation, land rent prices and population density are usually defined as functions of distance from the city center (Alonso, 1964; Mills, 1967; Fujita, 1989). More recently, in urban planning and transportation studies, particular attention has been given to the issue of accessiblity, as determined by the spatial distribution of potential destinations, its attractiveness and their ease of reach (Handy & Niemeier, 1997; Handy & Clifton, 2001).

Guided by urban economics and planning theory, a monocentric city would thus exhibit at its center higher access to goods and services (both in term of supply and ease of access), higher land use intensity and higher land prices, decreasing as one moves away from the center. Put another way, the closer to the city center, the higher the urbanization level. As such, for the purposes of this analysis urbanization level is conceptualized as a latent construct that accounts for the observed spatial distribution of the city in terms of supply of goods and services , land use intensity, transport mobility and land prices. Indicators were selected based on the results of an exploratory factor analysis (EFA) conducted on a set of potential indicators theoretically associated with urbanization levels. In addition, the spatial data used for this analysis (with the exception of population density) has the advantage of being available in the form of point data, which allows for a flexible definition of the analysis unit in order to address the MAUP issue discussed earlier. The four indicators used were:

A. Commercial Kernel density: Using location data of commercial facilities extracted from the geo-referenced phonebook data provided by ZENRIN Co., Ltd (2011), a Kernel density of all non-industrial services was estimated via ArcGIS, as a measure of supply of goods and services. As defined by Silverman (1986), the multivariate Kernel estimator can be written as

$$f(x)=\frac{1}{nh^2}\sum_{i=1}^{n} K\{\frac{1}{h}(\boldsymbol{x}-\mathbf{X}_i)\} \qquad (11)$$

where n is the sample size, h is the bandwidth or smoothing parameter, and K is a Kernel weighting function, defined for a bivariate variable **x** following Silverman (1986) as

$$K(x)=\begin{cases} 3\pi^{-1}(1-\mathrm{x}^{\top}\mathrm{x})^2 & \text{if } \mathrm{x}^{\top}\mathbf{x}<1 \\ 0 & \textit{otherwise} \end{cases} \qquad (12)$$

A symmetrical density function is drawn on each data point (each commercial facility) following the specified Kernel weighting function in equation (12) extending up to the defined bandwidth h at which point the weight becomes zero. The kernel density is thus the sum of these density values at each sampling point where the sampling mesh size was set at 50m x 50m.

Bandwidth *h* was defined rather arbitrarily at 500 meters. Nevertheless, estimated density values at bandwidths of 500 meters, 750 meters and 1,000 meters yielded high correlations, with all coefficients above 0.95. In that sense, since CFA aims at reproducing the observed variances and covariances of the data, the bandwidth specification is of little concern for the purposes of this analysis.

B. Population density: Population density was used as a measure of land use intensity. Since data from the 2005 national census was used (PASCO, 2005), at its finest resolution, the data is available only at the district level, as a result, it not possible to control for the zoning effect in the data.

C. Weekday transit frequency was used as a measure of transport mobility. Railway data was gathered from publicly available service timetables from each operator (Fukuoka City Transport Bureau, 2014; JR Kyushu, 2014; JR West, 2014; Nishi-Nippon Railroad Co., Ltd, 2014) while bus data was provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT, 2011a; MLIT, 2011b). Weekday transit frequencies for locations within 800 meters from train stations, and 300 meters from bus stops were calculated and added, resulting in a single transit accessibility index.

D. Land price: Land price data was provided by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT, 2013a; MLIT, 2013b) . Land prices were interpolated from 1,965 data points extracted from the combined datasets via *ArcGIS* using the nearest neighbor method.

4 Survey design and data characteristics

The main data source for this analysis was an online survey conducted in the city of Fukuoka City, Japan. The survey was conducted in December 2013, through a major net research company with over 2.3 million monitors all over Japan. The survey aimed at gathering four major types of information: (i) individual and household attributes, (ii) mobility biography (which includes relocation history and main modes of transport during different life stages (see Axhausen (2008)), (iii) attitudes related to transport and residential location, and (iv) travel behavior. The data gathered corresponds to a large extent to relevant covariates largely cited in the residential self-selection literature as playing in a role in co-explaining residential location and/or travel behavior (see Cao et al. (2009a) for an extensive review on the issue).

Regarding the sampling process, the target population was adult (over 20 years old) monitors living in Fukuoka City at the time of the survey. The sample size was set at 600 respondents, distributed by household composition following the 2010 population census of Japan. Monitors meeting the sampling criteria were sent a pre-survey to gather information on their household composition and verify whether the target cohort sample sizes can be met based on expected response rates. Respondents who completed the pre-survey were then randomly sampled based on the required cohort sample sizes and expected response rates, and asked to answer the main questionnaire. The final number of respondents of the main questionnaire was 656 persons, resulting in a response rate of 28.6% (Of the sampled subset of pre-survey respondents). The survey was pre-tested using a convenience sample of students and faculty in the Department of Urban Engineering of the University of Tokyo.

Table 1 compares the population distribution of Fukuoka City, to the sampling distribution. The single elder cohort was underrepresented in the sample by almost 7 percentage points; conversely, the single young cohort was over-represented the same amount.

Table 1: Individual and household sample characteristics

Household type	Frequency	Sample percentage	Population percentage
Single household	314	47.9%	47.7%
Of which: Young (age 20-64)	302	46.0%	39.2%
Of which: Elder (age 65 and over)	12	1.8%	8.5%
Couples only	101	15.4%	15.1%
Of which: Young (age 20-64)	60	9.1%	8.7%
Of which: Elder (age 65 and over)	41	6.3%	6.5%
Nuclear household (including single parent households)	201	30.6%	31.3%
Three generation household & others	40	6.1%	6.0%
Total	656	100%	100%

Population data source: 2010 population census of Japan

4.1 General characteristics of covariates

General sample characteristics were compared against the population characteristics taken mainly from the 2010 national census and the 2011 Private Income Statistical Survey (National Tax Agency, 2012) to check the representativeness of the sample. Due to space limitations, in addition to general socio-demographics, only covariates that made the final propensity score model (see Section 5.1.) are summarized in Table 2.

As is usual in online questionnaires, the average age in the sample is lower than the population sample suggesting a slight bias towards the young. Sample average household size is also larger, with a sample average of 2.21 against the population average of 2.01. Compared against the Private Income Statistical Survey for 2011 (National Tax Agency, 2012), in general the income distribution is rather similar to the national average distribution, although consistent with the web-survey literature (Couper, 2000), higher income households are slightly over-represented in samples while lower income cohorts are somewhat underrepresented.

In order to account for the effect of built environment characteristics at previous locations respondents were asked to indicate the address of the 3 places where they have spent most of their lives (besides their current location, which was asked separately). In addition, respondents were asked to state the life-course events, if any, motivating these relocations. The most frequently cited reasons for moving to the present location are employment-related reasons (19%) marriage (12%) and school-related reasons (10%).

Table 2: Individual and household sample characteristics

Variable name	Mean	Population mean	Std.Dev.
Household characteristics		2010 census data	
Household size	2.22	2.01	1.38
Number of children	0.46	-	0.82
Number of cars	0.70	0.98	0.67
Driver to car ratio	0.84	-	0.29
Number of workers	1.08	-	0.70
House is company/school lodge	0.03	-	-
Job located in city center	0.33	-	-
Household yearly income[1]		NTA National average	
Under JPY2,000,000	0.20	0.24	-
From JPY2,000,001 to JPY3,000,000	0.18	0.17	-
From JPY3,000,001 to JPY4,000,000	0.16	0.18	-
From JPY4,000,001 to JPY5,000,000	0.12	0.14	-
From JPY5,000,001 to JPY6,000,000	0.11	0.09	-
From JPY6,000,001 to JPY7,000,000	0.07	0.06	-
From JPY7,000,001 to JPY8,000,000	0.06	0.04	-
From JPY8,000,001 to JPY9,000,000	0.03	0.03	-
From JPY9,000,001 to JPY10,000,000	0.02	0.02	-
From JPY10,000,001 to JPY12,000,000	0.03	0.04	-
Over JPY12,000,000	0.02		-
Lifetime events motivating relocation			
Work (start, change)	0.19	-	-
School (enrollment, change)	0.12	-	-
Wedding	0.10	-	-
Empty nest	0.01	-	-
Job promotion	0.02	-	-
Individual characteristics		2010 census data	
Male	0.48	0.47	-
Age	43.43	48.64	13.39
Driver (Valid driver's license)	0.89	0.62	-
Worker (as primary occupation)	0.66	-	-
University degree holder	0.49	-	-
Attitudes and habits			
Attitude: Car lover	-0.02	-	0.99
Attitude: Urbanite	0.06	-	0.98
Car use Habit	4.18	-	3.37
Life ratio using transit as main travel mode	0.35	-	0.36
Log of weighted population density at previous locations	9.03	-	0.90

[1]JPY 1 = USD 0.091

Note: Car ownership and driver's license data gathered from the Kyushu Transport Bureau of the Ministry of Land Infrastructure, Transport and Tourism(MLIT), and the Northern Kyushu Comprehensive Travel Survey, respectively.

In terms of car ownership, the sample mean is estimated at 0.7 vehicles per household against a mean population value of 0.98 per household, the largest difference among measured variables. On the other hand, the ratio of driving license holders stands at 89% against a population ratio of 62%, although this difference might be partly explained by the exclusion of the under-20-years-old cohort.

Regarding attitudes and habits, automobile use habit was measured using the Response Frequency Index (RFI) proposed by Verplanken, Aarts, van Knippenberg, and van Knippenberg (1994) respondents were presented with 10 hypothetical trips and given six travel modes to choose from (See appendix for details on the measurement instrument). Habit was then measured as the simple summation of all the times car mode was selected. In terms of attitudes, a three factor Principal Component Analysis (PCA) was used to estimate the factors that explain unobserved attitudes towards residential location and transport. Respondents were asked to rate on a five point Likert Scale the level of agreement with 30 statements regarding private vehicles, public transport, non-motorized modes and residential location. The questionnaire design was largely based on previous studies by Kitamura et al. (1997) and Cao, Mokhtarian, and Handy (2009b), adapted to the Japanese case, and pre-tested accordingly. (See appendix for the CFA factor loading results).

4.2 Outcome variable of interest

The outcome variables considered for this analysis were home-based maintenance trip frequencies by mode. Maintenance activities refer to those activities other than subsistence activities (work and school related activities) that need to be conducted in the course of daily life such as grocery shopping, visits to the doctor, going to the bank, and other personal business. Discretionary activities were excluded as discretionary activity generation might be more dependent on factors such as social network characteristics, which are not controlled for in the current dataset. Respondents were asked to state the number of trips (excluding the return trip) taken during the week before up to the survey day by purpose and mode (see Table 3).

Table 3: Summary of reported travel behavior characteristics of the sample

Variable name	Mean	Std. Dev.	Minimum	Maximum
Total home-based maintenance trips	4.358	3.616	0	50
Of which: Car trip	1.321	1.955	0	11
Of which: Transit trips	0.295	0.894	0	10
Of which: Non-motorized trips	2.741	3.301	0	40

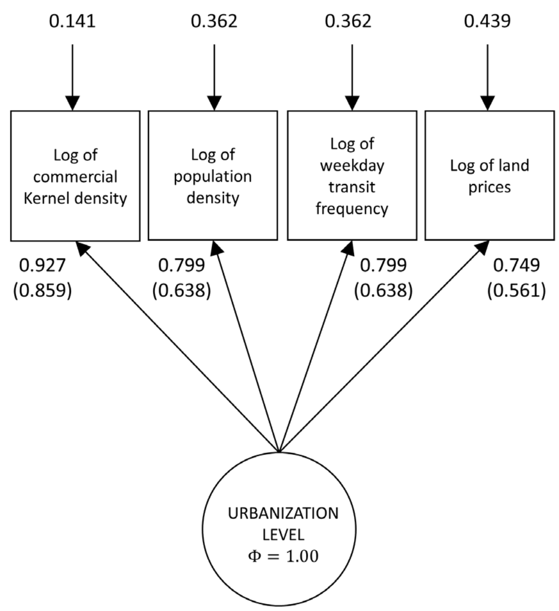
5 Model specification and results

5.1 Urbanization index model

Following the explanation provided in Section 3.4., The CFA model was estimated using MPLUS 6, developed by Muthen and Muthen (2010). Units were excluded from the analysis if (i) the population density at any given unit is equal to zero, or (ii) data for any of the indicator variables is not available for a given unit. This yielded an effective sample size of 18,485 cells out of the total 19,686 cells in which the study area was tessellated.

As a result of the multivariate non-normality condition of the indicator variables the robust maximum likelihood estimator was used. Although the issue of goodness of fit statistics remains still a hotly debated subject (Marsh, Hau, & Wen, 2004; Saris, Satorra, & Van der Veld, 2009; Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011) Goodness of fit acceptable thresholds are guided by the values recommended by Hu and Bentler (1999) as follows: Standardized root mean square residual SRMR (≤0.08), comparative fit index CFI (≥0.95), Tucker-Lewis index TLI (≥0.95), and a root mean square error of approximation (RMSEA) cut-off value of ≤ 0.05.

With 2 degrees of freedom, the Chi-square statistic is significant at the 0.01 level. This might suggest that the model does not reproduce the observed variances and covariances of the indicators well enough; nevertheless, Chi-square is inflated by sample size, thus tending to routinely reject large sample size solutions (Brown, 2006). Other indices not sensitive to sample size, however, suggest an acceptable model fit. RMSEA is 0.037, with a confidence interval of 0.028 and 0.046 at its lower and upper boundaries respectively. CFI and TLI are 0.999 and 0.996 respectively, while the standardized root mean square residual (SRMR) is 0.005. The path diagram of the estimated latent variable is shown in Figure 1.



Chi-Square test of model fit (d.f.) 51.38 (2); p-value: 0.000; RMSEA (C.I. 90%) : 0.037 (0.028, 0.046)
Probability RMSEA ≤.05 : 0.994; CFI: 0.999; TLI: 0.996; SRMR: 0.005
Value in parenthesis is total explained variance by the factor.
All parameter estimates are significant at the p < 0.01 level.
Due to multivariate non-normality, estimator is Robust Maximum Likelihood.

Figure 1: Path diagram of "Urbanization Level" latent variable

Another criteria for evaluating the model was the modification indices, presented in Table 4. Modification indices reflect Chi-square changes given freely estimating the error covariances. In practice, modification indices above the 3.84 level suggest areas of strain in the model or potential improvements. However; since the indices reflect changes in Chi-square, they are also sensitive to large sample sizes. Fit-improving specification search guided by a sound theoretical reasoning is a widely accepted practice in the CFA field, and given the complexity of spatial dynamics, arguments can be put forth to support this approach. That is, the theory that other sources of covariation other than the urbanization latent factor exist among indicators is not at all unrealistic. However, in the absence of a well-established error theory to guide these specifications the current more parsimonious model was selected with error measures (unique variances) assumed random.

Table 4: CFA model modification indices

With statements	Modification index	E.P.C.	STD E.P.C.
Log of population density with log of Kernel density	9.714	-0.069	-0.068
Log of transit frequency with log of Kernel density	19.278	-0.105	-0.095
Log of transit frequency with log of population density	51.760	0.144	0.081
Log of land price with log of Kernel density	51.744	0.048	0.127
Log of land price with log of population density	19.230	-0.026	-0.044
Log of land price with log of transit frequency	9.714	-0.020	-0.031

E.P.C.: Expected parameter change; STD E.P.C.: Fully standardized expected parameter change
Only indices above 3.84 are reported

All estimated parameters were statistically significant at the $p < 0.01$ level. Factor loadings suggest that all indicators are strongly related with the latent factor urbanization level, especially the log of commercial density, whose total explained variance stands at 85.9%. Figure 2 illustrates the spatial distribution of the estimated urbanization level latent variable. Clearly, there is a marked mono-centricity in the spatial distribution of the city, with the highest levels of urbanization concentrated mainly around Chuo ward and spreading outwards.

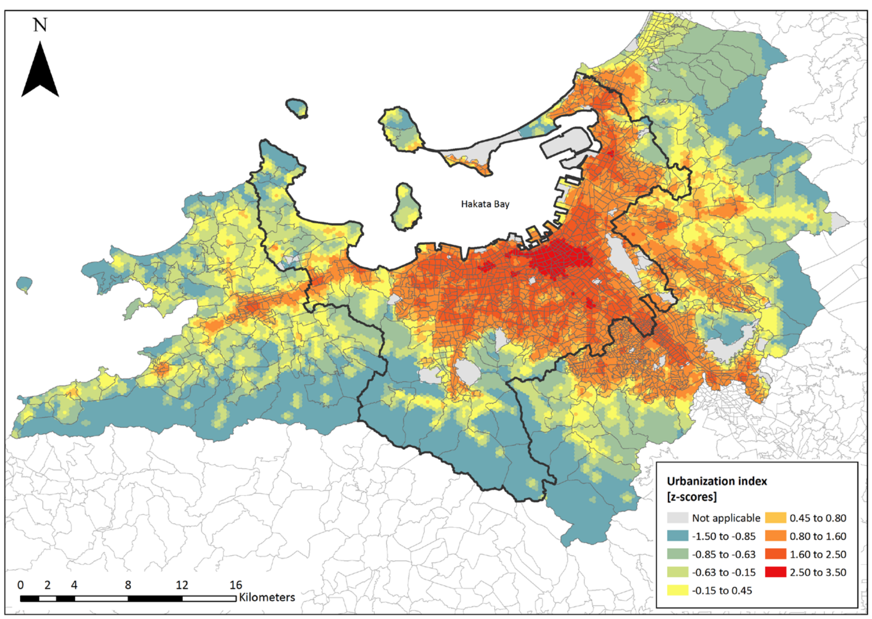


Figure 2: Urbanization level map of Fukuoka City

5.2 Estimating the propensity score function

As explained in Section 3.2, an estimate of the propensity score function $\hat{\theta}$ for the continuous treatment variable urbanization level is estimated through an OLS regression. Covariate selection was based both on findings from the literature as well as the theoretical considerations. Three types of variables are included in the regression function: household characteristics, lifetime events motivating the relocation and individual characteristics such as education level, habits and attitudes, which are assumed representative of those members involved in the residential location choice decision. Estimation results are presented in Table 5. R-squared of the final model was 0.25 suggesting an acceptable model fit. Note that the propensity score function is the same for both the simulations and the empirical analysis.

It is important to note that as a prediction model, the object of interest of this regression is not the individual coefficients of each explanatory variable, but the scalar estimate $\hat{\theta}$. Following the balancing score assumption described in equation (1), $\hat{\theta}$ balances all the covariates thought to affect treatment allocation. This warrants the inclusion in the final model of variables that although theoretically significant might be rendered insignificant or exhibit the wrong sign due to high correlations among covariates.

Table 5: Propensity Score OLS Estimation Results

N	491	S.E. of Regression	0.5331
Parameters	19	R-square	0.25
d.f.	472	Adj. R-square	0.22
RSS	134.14	F test (p-value)	8.66 (.0000)
Variable	β	S.E.	t-Stat
Constant	1.505	0.337	4.467
Household characteristics			
Household size	-0.087	0.039	-2.219
Number of children	0.110	0.053	2.079
Number of cars	-0.164	0.060	-2.726
Driver to car ratio	0.249	0.100	2.477
Number of workers	0.049	0.037	1.339
High Income (Over JPY 10,000,000)[1]	0.141	0.066	2.144
House is company/school lodge	-0.193	0.132	-1.465
Job located in city center	0.072	0.048	1.487
Lifetime events motivating relocation			
School (Start, change)	0.132	0.080	1.648
Wedding	-0.156	0.079	-1.981
Empty nest	0.707	0.327	2.161
Job promotion	-0.201	0.149	-1.354
Individual characteristics			
University degree holder	0.060	0.047	1.258
Attitudes and habits			
Attitude: Car lover	-0.035	0.025	-1.392
Attitude: Urbanite	0.059	0.025	2.368
Car use Habit	-0.034	0.012	-2.796
Life ratio using transit	0.103	0.068	1.503
Log of weighted population density at previous locations	0.049	0.033	1.517

[1]JPY 1= USD 0.091

To verify the balancedness of covariates given $\hat{\theta}$, as suggested by Imai and Van Dyk (2004) each covariate was regressed against the original treatment variable. The same regressions were then run a second time but this time conditioning on $\hat{\theta}$. OLS was used for continuous covariates while binary logit was used for dummy covariates. As Figure 3 illustrates, without controlling for $\hat{\theta}$, most covariates are strongly correlated with the treatment, but once conditioned on the propensity score estimate, this correlation is considerably reduced, evident in the drop of the t-statistics for each covariate.

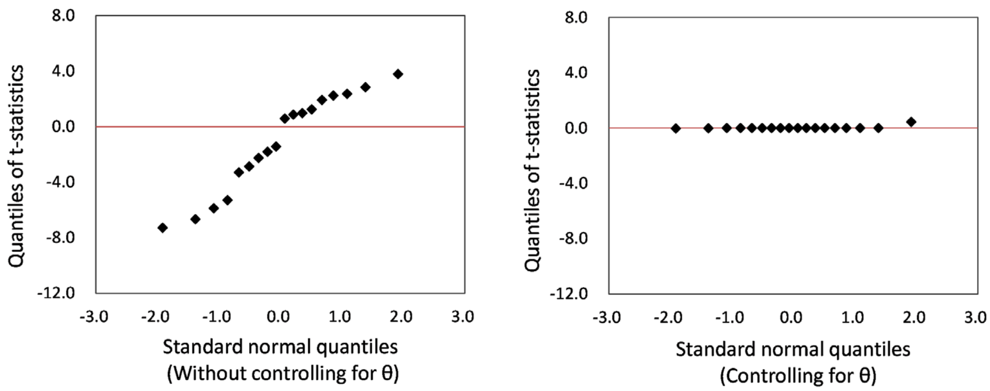


Figure 3: Standard Normal Quantile Plots of t-Statistics of covariates with and without controlling for the propensity score estimate

5.3 Measuring the performance of the propensity score stratification against OLS

As discussed in Section 3.3, for each simulation, treatment effect is estimated using a full-covariate OLS, and propensity score stratification stratified on $\hat{\theta}$ into roughly equal sub-classes *J*, where *J*= 3, 5 and 7 strata respectively. In addition all propensity score models are estimated with no covariates, and with the full set of covariates.

The performance of each model is compared against the full-covariate OLS estimates (statistical control approach), measured in terms of absolute bias where

$$\widehat{ABias}=\frac{1}{R}\sum_{r=1}^{R}\hat{\delta}-\delta \tag{13}$$

and mean squared error where

$$\widehat{MSE}=\frac{1}{R}\sum_{r=1}^{R}(\hat{\delta}-\delta)^2 \tag{14}$$

where $\hat{\delta}$ is the estimated treatment effect and R is the number of replications.

In terms of treatment effects, performance comparison is conducted first under the assumption of a fixed treatment effect that is constant to all individuals, and second, under the assumption of a variable treatment effect defined as a function of another variable. The constant treatment effect parameters used in the simulations were the estimated OLS values from full covariate models on the real dataset. In the variable treatment case the treatment parameters were defined as a function of car use habit, where for individual *i*

$$\tilde{\delta}_i=10^{-1}\,(10\text{-}H)\,\delta_m \tag{15}$$

where H is the car use habit index as measured by the Response Frequency Index method, and δ_m is equivalent to the constant treatment parameter for mode m. Under this function, the treatment effect tends to zero as the car use habit increases. This is, however, an arbitrary function in order to illustrate the variable treatment case, but another function might have been used as well.

Simulated results are shown in Tables 6 and 7, for car trips and NMM trips respectively. Results are given in percentage bias change (or MSE change) relative to the full covariate OLS estimates. Positive values indicate that the model underperforms the benchmark OLS model (bias increases relative to OLS), while negative values suggest that the model outperforms the benchmark model (bias decreases relative to OLS).

Results suggest that performance of the propensity score model against OLS is dependent on (i) the inclusion or not of covariates, and (ii) the linearity level of the model. Full-covariate propensity score models in general outperform the no-covariate models. Although in a very few cases the no-covariate models outperformed all other models, more than 50% of these models underperformed against the benchmark models, which supports the inclusion of all covariates in the estimation, a point that has also been noted by Imai and van Dyk (2004).

In terms of linearity levels, for all levels except the extremely non-linear case (R^2 0.10 to 0.15) the propensity score model generally outperforms the benchmark models with reductions in absolute bias and minimum square error of 2-37% and 4-78%, respectively. In the case of the 7-strata all-covariate models, two instances of bias increase were observed (1.2% and 5.6%). Even in those cases reductions of 21.7% and 33.5% were observed in mean squared error, which weighs heavier on the efficiency of the estimate relative to absolute bias.

In the case of extreme non-linearity; however, the propensity score models yielded mixed results, in some cases severely underperforming against the benchmark models, suggesting poor reliability at such levels of non-*linearity*.

Table 6: Performance of the propensity score approach against OLS for car trips

Constant treatment	**3 Strata**		**5 Strata**		**7 Strata**	
Difference in absolute bias	N.C.	A.C.	N.C.	A.C.	N.C.	A.C.
Extremely non-linear	8.17%	-7.58%	12.50%	-4.30%	20.04%	79.84%
Highly non-linear	6.63%	-7.09%	3.48%	-7.94%	3.03%	1.22%
Non-linear	12.11%	-10.74%	3.48%	-12.16%	4.93%	-8.28%
Moderately non-linear	14.67%	-15.58%	3.23%	-28.32%	0.78%	-29.70%
Difference in MSE						
Extremely non-linear	16.74%	-20.32%	26.24%	-15.23%	46.70%	214.23%
Highly non-linear	91.90%	-10.90%	69.52%	-78.00%	42.39%	-49.81%
Non-linear	15.54%	-18.41%	3.73%	-23.85%	4.67%	-21.65%
Moderately non-linear	30.75%	-22.98%	5.53%	-45.38%	-0.77%	-46.82%
Variable treatment	**3 Strata**		**5 Strata**		**7 Strata**	
Difference in absolute bias	N.C.	A.C.	N.C.	A.C.	N.C.	A.C.
Extremely non-linear	-25.13%	12.99%	-27.88%	-6.97%	-14.46%	29.79%
Highly non-linear	-2.00%	-5.43%	-6.07%	-3.68%	-5.00%	5.59%
Non-linear	5.02%	-14.04%	-4.47%	-19.45%	-3.46%	-15.88%
Moderately non-linear	32.55%	-6.80%	6.28%	-23.80%	-0.96%	-24.47%
Difference in MSE						
Extremely non-linear	-43.94%	27.67%	-47.99%	-13.46%	-26.83%	68.45%
Highly non-linear	-13.06%	-8.80%	-13.31%	-8.98%	-13.69%	-4.54%
Non-linear	-1.93%	-21.76%	-11.70%	-32.15%	-11.00%	-33.48%
Moderately non-linear	61.65%	-9.91%	8.28%	-39.89%	-5.92%	-40.92%

N.C. No Covariates: A.C. All covariates

Table 7: Performance of the propensity score approach against OLS for non-motorized trips

Constant treatment	**3 Strata**		**5 Strata**		**7 Strata**	
Difference in absolute bias	N.C.	A.C.	N.C.	A.C.	N.C.	A.C.
Extremely non-linear	-37.43%	42.62%	-10.55%	139.41%	-22.34%	57.81%
Highly non-linear	11.19%	-4.09%	7.94%	-2.50%	4.96%	0.00%
Non-linear	21.54%	-8.48%	9.16%	-12.70%	4.92%	-9.42%
Moderately non-linear	9.43%	-14.55%	-0.32%	-29.29%	-1.17%	-31.46%
Difference in MSE						
Extremely non-linear	-73.18%	129.48%	-35.78%	567.76%	-54.90%	162.05%
Highly non-linear	75.19%	-20.85%	72.89%	-20.36%	60.32%	-37.63%
Non-linear	40.50%	-14.68%	17.18%	-24.42%	8.56%	-21.19%
Moderately non-linear	24.98%	-22.13%	1.12%	-46.04%	-3.64%	-49.91%
Variable treatment	**3 Strata**		**5 Strata**		**7 Strata**	
Difference in absolute bias	N.C.	A.C.	N.C.	A.C.	N.C.	A.C.
Extremely non-linear	-98.11%	-7.97%	-82.41%	20.47%	-60.08%	-37.41%
Highly non-linear	9.39%	-5.78%	8.77%	-5.13%	6.31%	-6.38%
Non-linear	12.52%	-7.47%	5.51%	-11.70%	2.61%	-11.38%
Moderately non-linear	5.98%	-14.02%	-1.53%	-21.02%	0.63%	-22.40%
Difference in MSE						
Extremely non-linear	-99.97%	-15.44%	-96.94%	44.65%	-84.20%	-60.96%
Highly non-linear	73.97%	-8.85%	54.26%	-49.70%	21.55%	-57.65%
Non-linear	21.92%	-11.40%	10.27%	-22.42%	4.32%	-23.54%
Moderately non-linear	17.09%	-22.53%	-1.54%	-35.93%	-1.87%	-38.90%

N.C. No Covariates: A.C. All covariates

5.4 Empirical application to home-based maintenance trips

Having demonstrated the bias reduction potential of the propensity score approach under certain conditions, an empirical analysis is conducted using the Fukuoka City dataset described in Section 4, the urbanization level latent index, and the propensity scores estimated in sections 5.1 and 5.2 respectively. In addition, based on the modifiable areal unit problem described earlier, a multi-scale analysis is conducted to evaluate the robustness of the treatment effect estimates to changes in scale.

Although given the way the treatment variable was estimated, both the zoning and scale problems are to some extent controlled for. However, the optimal scale of analysis, that is, the actual spatial scale that households consider when evaluating residential location alternatives is in practice not known. Guo and Bhat (2007) addressed this issue in terms of residential location choice models by operationalizing several definitions of "neighborhoods". In addition to the census tracts, Guo and Bhat analyzed radial neighborhoods and network band models given different radii, namely, 0.4 km, 1.6 km and 3.2 km from each residential location alternative. Since the improvement of the more complex network band neighborhood was rather marginal, for this study the simpler radial network operationalization is used.

As illustrated in Figure 4, the first scale of analysis (Scale 1) is the same scale at which the urbanization level index was estimated, that is, a 300m diameter hexagon. The second and third scales take the unweighted average of the urbanization level of all units within a 1500 meter and 3000 meter radii respectively. In addition to the radial neighborhood operationalization, a more conceptually appealing analysis scale is proposed. The fourth scale of analysis assigns a weight to surroundings areas as a function of distance from each unit centroid via a kernel density function as described in Section 4.2 so that closer locations are given more importance than more distant ones. Recall that the kernel density function is rather insensitive to bandwidth (radius) specification, making the radius specification irrelevant.

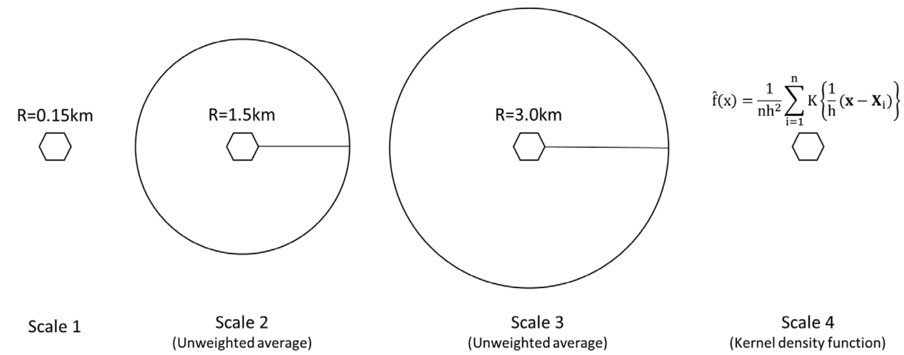


Figure 4: Diagram of scale definitions for multi-scale analysis

Tables 8 and 9 summarize the urbanization level average treatment effect estimates for full-covariate OLS against full-covariate 5-strata and 7-strata propensity score models at each spatial scale respectively. For all models, at any scale the direction of the effects is as hypothesized, negative for car trips and positive for non-motorized modes, thus supporting the idea of a mode substitution mechanism between car and non-motorized trips given changes in urbanization level. It is important to note that the R^2 of the OLS models were around 0.25 and 0.45 for the non-motorized and car models respectively, putting them within the range where the propensity score approach was shown to outperform the benchmark models.

Table 8: OLS and 5 strata propensity score estimates of urbanization level effect on home-based maintenance trips at different scales (Full-covariate models)

		Scale 1		**Scale 2**		**Scale 3**		**Scale 4**	
Model		OLS	5 Strata	OLS	5 Strata	OLS	5 Strata	OLS	5 Strata
Car trip frequency model	β	-0.201	-0.200	-0.145	-0.217	-0.127	-0.178	-0.131	-0.217
	t-Stat	-4.794	-3.381	-3.191	-5.020	-2.477	-4.106	-3.273	-5.110
	%Δ	-0.1%		50.0%		39.5%		65.7%	
NMM trip frequency model	β	0.151	0.152	0.125	0.156	0.089	0.179	0.103	0.177
	t-Stat	2.595	2.604	1.924	2.710	1.215	3.230	1.746	3.025
	%Δ	0.4%		24.8%		101.0%		71.9%	

Table 9: OLS and 7 strata propensity score estimates of urbanization level effect on home-based maintenance trips at different scales (Full-covariate models)

		Scale 1		Scale 2		Scale 3		Scale 4	
Model		OLS	7 Strata	OLS	7 Strata	OLS	7 Strata	OLS	7 Strata
Car trip frequency model	β	-0.201	-0.196	-0.145	-0.223	-0.127	-0.205	-0.131	-0.217
	t-Stat	-4.794	-4.326	-3.191	-4.592	-2.477	-4.220	-3.273	-4.381
	%Δ	-2.4%		54.1%		61.0%		65.8%	
NMM trip frequency model	β	0.151	0.160	0.125	0.181	0.089	0.172	0.103	0.187
	t-Stat	2.595	2.545	1.924	2.989	1.215	3.023	1.746	3.245
	%Δ	5.9%		45.3%		92.4%		81.7%	

At Scale 1, OLS and propensity score treatment effect estimates are rather similar, with differences ranging from 0.4% to 6% However, at different spatial scales, the propensity score estimates are more robust, while the OLS estimates deteriorate quickly with difference in estimates up to 101%. Furthermore, in the NMM case, the t-statistics for the OLS estimates fall below the 5% threshold for all but the Scale 1 estimates, becoming insignificant at any significance level for the Scale 3 estimates. As stated earlier, the optimal scale of analysis is in practice unknown; however, results suggests that the propensity score estimates are more stable and robust to changes in analysis scale.

6 Discussion and conclusion

This study validated through Monte Carlo simulation the propensity score approach as a tool to examine the connection between the built environment and travel behavior from a cross-sectional approach. It is shown that under the ignorability of treatment assumption, the causal effect of urbanization level on travel behavior can be estimated. Stratification on the propensity score was shown to outperform the benchmark OLS models with maximum reductions in absolute bias and minimum square error of 37% and 78%, respectively. However, at extreme levels of non-linearity the propensity score models yielded mixed results, suggesting low reliability.

As discussed in earlier, a continuous urbanization level treatment, as the one used here allows for a more precise understanding of the built environment effect on travel behavior at all levels of the urbanization spectrum without the need to arbitrarily draw a defining line between "urban" and "suburban" which binary treatment models might be highly sensitive to. Empirical analysis of data also suggested that the propensity score approach is more robust to changes in the scale of analysis.

In terms of the propensity score function, the importance of the strong ignorability of treatment assumption cannot be over-emphasized. That is, the assumption that the distribution of treatment outcomes are independent from the distribution of treatment assignment given the propensity score is crucial to the unbiasedness of estimates. Nevertheless, in practice it is impossible to know how well the estimated function approximates the true population function. In order to estimate the propensity score function, relevant variables largely cited in the literature introduced in the model, hence, it is assumed at the estimated function is a good estimate of the true unknown function. However, the risk of misspecification is certainly non-trivial. In that sense, much care should be placed in estimating the propensity score function, as much of the validity of the analysis depends on it.

The main travel behavior dimension analyzed in this study relate to trip frequencies by mode. However, other relevant dimensions should be analyzed to strengthen the conclusions presented in this article. Certainly the propensity score approach presented here can be used to analyze continuous variables such as travel distance, or fuel consumption, provided reliable data is available.

In general, findings support the notion that the built environment has a significant effect on travel behavior, specifically, on trip frequency by mode, providing some empirical evidence to the claims of compact city advocates. Nevertheless, it is important to note that the issue at hand is more complex that just retrofitting or promoting a certain re-development model. In spite of the existence of a causal relation, residential location not only is a self-selecting process guided by household life-stage, lifestyle and preferences, but it's at the same time constrained by the supply and demand dynamics of the real estate market. In that sense, a mismatch between supply and demand might hamper efforts to promote compact city paradigms. Even for households that wish to move to the city center, rent costs might be prohibitively expensive, pushing households to more suburban areas where they can afford more space. In the case of Japanese cities, this problem is extenuated by lax urban control laws that allow development to expand even beyond the so-called Urban Control Areas, thus promoting suburbanization, perhaps unintentionally.

7 Acknowledgments

This study was supported by JSPS KAKENHI Grant No. 23246091 and The University of Tokyo's Excellent Graduate Schools (EGS) Support Fund for Young Researchers Grant No. T10027. All spatial data used for the analysis presented in this article were provided by the Center for Spatial Information Science of The University of Tokyo. CSIS joint research No.479.

References

Alonso, W. (1964). *Location and land use: Towards a general theory of land rent.* Cambridge, MA: Harvard University Press.

Axhausen, K. (2008). Social networks, mobility biographies, and travel: Survey challenges. *Environment and Planning B: Planning and Design, 35*(6), 1–17.

Birch, C. P., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling, 206,* 347–359.

Boarnet, M., & Sarmiento, S. (1998). Can land-use policy really affect travel behavior? A study of the link between non-work travel and land-use characteristics. *Urban Studies, 35*(7), 1155–1169.

Boer, R., Zheng, Y., Overton, A., Ridgeway, G. K., & Cohen, D. A. (2007). Neighborhood design and walking trips in ten U.S. metropolitan areas. *American Journal of Preventive Medicine, 32*(4), 298–304.

Bohte, W., Maat, K., & van Wee, B. (2009). Measuring attitudes in research on residential self-selection and travel behavior: A review of theories and empirical research. *Transport Reviews: A Transnational Transdisciplinary Journal, 29*(3), 325–357.

Brown, T. (2006). *Confirmatory factor analysis for applied research.* New York: The Guilford Press.

Cao, X. (2010). Exploring causal effects of neighborhood type on walking behavior using stratification of propensity score. *Environment and Planning A, 42*, 487–504.

Cao, X., Handy, S., & Mokhtarian, P. (2006). The influences of the built environment and residential self-selection on pedestrian behavior: Evidence from Austin TX. *Transportation, 33,* 1–20.

Cao, X., Mokhtarian, P., & Handy, S. (2009a). Examining the impacts of residential self-selection on travel behavior: A focus on empirical findings. *Transport Reviews, 29*(3), 359–395.

Cao, X., Mokhtarian, P., & Handy, S. (2009b). The relationship between the built environment and nonwork travel: A case study of Northern Carolina. *Transportation Research Part A, 43,* 548–559.

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Do changes in neighborhood lead to changes in travel behavior? A structural equations modeling approach. *Transportation, 34,* 535–556.

Cao, X., Yu, Z., & Fan, Y. (2010). Exploring the connections among residential location, self-selection, and driving: Propensity score matching with multiple treatments. *Transportation Research Part A, 44,* 797–805.

Chatman, D. G. (2009). Residential choice, the built environment, and nonwork travel: Evidence using new data and methods. *Environment and Planning A, 41*, 1072–1089.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *The Public Opinion Quarterly, 64*(4), 464–494.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography, 31*(5), 471–479.

Fotheringham, A. S., & Wong, D. W. S. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A, 23*(7), 1025–1044.

Fujita, M. (1989). *Urban economic theory: Land use and city size.* Cambridge, England: Cambridge University Press.

Fukuoka City Transport Bureau. (2014). 福岡市交通局・時刻表 [Fukuoka City Transport Bureau: Timetable]. Retrieved from: http://subway.city.fukuoka.lg.jp/schedule/

Guo, J., & Bhat, C. (2007). Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography, 15,* 31–45.

Handy, S., Cao, X., & Mokhtarian, P. (2005). Correlation or causality between the built environment and travel behavior? Evidence from Northern Carolina. *Transportation Research Part D, 10,* 427–444.

Handy, S., Cao, X., & Mokhtarian, P. (2006). Self-selection in the relationship between the built environment and walking. *Journal of the American Planning Association, 72*(1), 55–74.

Handy, S. L., & Clifton, K. J. (2001). Evaluating neighborhood accessibility: Possibilities and practicalities. *Journal of Transportation Statistics*, *4*(2-3), 67–78.

Handy, S. L., & Niemeier, D. A. (1997). Measuring accessibility: An exploration of issues and alternatives. *Environment and Planning A, 29*(7), 1175–1194.

Heckman, J., Ichimura, H., Smith, J., & Todd, P. (1998). *Characterizing selection bias using experimental data.* Cambridge, MA: National Bureau of Economic Research.

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods, 16(*3), 349–336.

Horvitz, D., & Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association, 47,* 663–685.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Imai, K., & van Dyk, D. A. (2004). Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association, 99*(467), 854–866.

Imbens, G. M., & Wooldridge, J. M. (2008). *Recent developments in the econometrics of program evaluation,* Cambridge, MA: National Bureau of Economic Research.

Jelisnki , D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology, 11*(3), 129–140.

JR Kyushu. (2014). JR九州駅時刻表 [JR Kyushu station timetables]. Retrieved from http://www.jrkyushu-timetable.jp/jr-k_time/map_fukuoka.html

JR West. (2014). 鉄道のご案内　JRおでかけネット [Train information JR Odekake Net]. [Online] Retrieved from https://www.jr-odekake.net/railroad/#eki

Khattak, A. J., & Rodriguez, D. (2005). Travel behavior in neo-traditional neighborhood developments: A case study in USA. *Transportation Research Part A, 39*, 481–500.

Kitamura, R., Mokhtarian, P., & Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation, 24,* 125–158.

Krizek, K. J. (2003). Operationalizing neighborhood accessibility for land use travel behavior research. *Journal of Planning Education and Research, 22,* 270–287.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal, 11*(3), 320–341.

Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *The American Economic Review, 57*(2), 197–210.

MLIT. (2011a). 国土数値情報バスルートデータ [National land numerical information: Bus route data]. Ministry of Land, Infrastructure, Transport and Tourism. Retrived from http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N07.html

MLIT. (2011b). 国土数値情報バス停留所データ [National land numerical information: Bus stop data]. Ministry of Land, Infrastructure, Transport and Tourism. Retrieved from http://nlftp.mlit.go.jp/ksj/jpgis/datalist/KsjTmplt-P11.html

MLIT. (2013a). 国土数値情報　地価公示調査 [National land numerical information: Published land price survey]. Ministry of Land, Infrastructure, Transport and Tourism. Retrieved from http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L01-v2_3.html

MLIT. (2013b). 国土数値情報 都道府県地価調査 [National land numerical information: National survey on land]. Ministry of Land, Infrastructure, Transport and Tourism. Retrieved from http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-L02-v2_3.html

Mokhtarian, P., & Cao, X. (2008). Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research B, 42*(3), 204–228.

Muthen, L., & Muthen, B. (2010). *Mplus user's guide.* 7th ed. Los Angeles: Muthen & Muthen.

Naess, P. (2009). Residential self-selection and appropriate control variables in land use: Travel studies. *Transportation Reviews: A Transnational Transdisciplinary Journal, 3*, 293–324.

Naess, P. (2014). Tempest in a teapot: The exaggerated problem of transport-related residential self-selection as a source of error in empirical studies. *Journal of Transport and Land Use, 7*(3), 57–79.

National Tax Agency. (2012). Heisei 23 nenbun Minkan kyuuyo jittai chousa [Private income statistical survey for 2011]. Tokyo, Japan. Retrieved from http://www.nta.go.jp/kohyo/tokei/kokuzeicho/minkan2011/minkan.htm

Nishi-Nippon Railroad Co., Ltd. (2014). にしてつ時刻表 [Nishitetsu Timetable]. Retrieved from http://jik.nishitetsu.jp/

PASCO. (2005). 国勢調査地図データ 統計地図 [National census statistical and map data]. PASCO Corporation.

Putman, S. H., & Chung, S. (1989). Effect of spatial system design on spatial interaction models: The spatial definition problem. *Environment and Planning A, 21,* 27–46.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rosenbaum, P., & Rubin, D. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*(387), 516–524.

Rubin, D. B., & Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association, 95*(450), 573–585.

Saris, W. E., Satorra, A., & Van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 561–582.

Scheiner, J., & Holz-Rau, C. (2013). Changes in travel mode use after residential relocation: A contribution to mobility biographies. *Transportation, 40,* 431–458.

Scheiner, J. (2014). Residential self-selection in travel behavior: Toward an integration into mobility biographies. *Journal of Transport and Land Use, 7*(3), 15–28.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* New York: Chapman and Hall.

Troncoso Parady, G., Chikaraishi, M., Takami, K., Ohmori, N., & Harata, N. (2015). On the effect of the built environment and preferences on non-work travel: Evidence from Japan. *European Journal of Transportation and Infrastructure Research, 15(*1), 55–65.

Troncoso Parady, G., Takami, K., & Harata N. (2014a). Connection between built environment and travel behavior: Propensity score approach under a continuous treatment regime. Transportation Research Record, *2453*(1), 137–144.

Troncoso Parady, G., Chikaraishi, M., Takami, K., & Harata, N. (2014b). A panel approach to understanding the effect of the built environment on travel behavior: A case study of the Kashiwanoha Area, Chiba, Japan. *Urban and Regional Planning Review, 1*, 18–38.

Vance, C., & Hedel, R. (2007). The impact of urban form on automobile travel: Disentangling causation from correlation. *Transportation*, *34,* 575–588.

Verplanken, B., Aarts, H., van Knippenberg, A., & van Knippenberg, C. (1994). Attitude versus general habit: Antecedents of travel mode choice. *Journal of Applied Social Psychology, 24*(11), 285–300.

ZENRIN Co., Ltd. (2011). テレデータPack！[Telepoint Pack DB]. ZENRIN Co., Ltd: s.n. http://www.zenrin.co.jp/product/gis/teldata/teldt.html

Zhang, M., & Kukadia, N. (2005). Metrics of urban form and the modifiable areal unit problem. *Transportation Research Record, 1902,* 71–79.

8 Appendix

Table A1: Response frequency index measurement instrument

	Car	Transit	Bicycle	Walk	Motor-bike	Other
Go to the super market						
Meet a friend who lives out of town						
Go to the convenience store						
Go to the beach						
Go to a sky resort						
Go to the park						
Go to a hot spring						
Go to the movies						
Have dinner at a restaurant						
Go buy clothes						

Table A2: PCA Factor loadings for residential location and transport related attitudes

Factor	Statement	Loading
Car dependent & Suburban preference	I prefer living in a place where it's easy to guarantee a parking space	0.643
	Driving a car gives me a sense of freedom	0.633
	I like driving	0.609
	More roads should be built to improve traffic conditions	0.589
	In general, the car is the safest transport mode	0.579
	I prefer living in a large house over having good transit accessibility	0.569
	Owning a car is a symbol of social status	0.551
	The suburbs are a better places to raise a family than the city center	0.540
	Before giving up driving altogether, I'd switch to an environmentally friendly car	0.465
	As much as possible, I prefer not living in multi-family housing	0.388
	I prefer living in a place close to a large-scale shopping center	0.368
	The cost of riding public transport are higher than the costs of driving a car	0.359
	Buses and trains are unreliable	0.335
Pro-alternative modes	Whenever possible, I prefer riding public transport than driving	0.745
	Whenever possible, I prefer walking or riding a bicycle than driving	0.744
	I like walking	0.630
	I like riding a bicycle	0.538
	Riding the bus or the train is comfortable	0.519
	Gas prices should be raised as a countermeasure to traffic jams and air pollution	0.485
	Using tax money to pay for public transport improvements is a good investment	0.466
Urbanite	I prefer living in a place with good access to the city center	0.789
	Living in a place with good transit accessibility is important	0.770
	If I were to move, I would prefer moving to the city center than to the suburbs	0.705
	Living in walking or biking range of different shops is important	0.691
	I prefer living in a place that is close to different amenities, even if I have to live in a smaller house	0.660
	I prefer living in a large house over having good transit accessibility	-0.372
	As much as possible, I prefer not living in multi-family housing	-0.328

Extraction Method: Principal Component Analysis; Rotation Method: Promax with Kaiser Normalization.