Impact of Built Environment on First- and Last-Mile Travel Mode Choice

Citation	Mo, Baichuan et al. "Impact of Built Environment on First- and Last-Mile Travel Mode Choice." Transportation Research Record: Journal of the Transportation Research Board (July 2018): 036119811878842. © 2018 National Academy of Sciences: Transportation Research Board
As Published	http://dx.doi.org/10.1177/0361198118788423
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/120463
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/4.0/

Impact of Built Environment on First- and Last-Mile Travel Mode Choice

Baichuan Mo
Department of Civil Engineering
Tsinghua University
Address: Tsinghua University, Beijing, China 100084
Phone: +86-151-1002-6340
Email: mbc14@mails.tsinghua.edu.cn

Yu Shen, PhD (Corresponding Author)
Future Urban Mobility IRG
Singapore-MIT Alliance for Research and Technology
Address: 1 CREATE Way, Singapore 138602
Phone: +65-6601-1231
Email: yushen@smart.mit.edu

Jinhua Zhao, PhD
Department of Urban Studies and Planning
Massachusetts Institute of Technology
Address: 77 Massachusetts Ave, Cambridge, MA 02139
Phone: +1-617-324-7594
Email: jinhua@mit.edu

The total number of words is 7,417 (6,167 words + 2 figures + 3 tables).
Submitted to the TRB 97th Annual Meeting, Nov 15th, 2017

ABSTRACT

The paper studies the impacts of built environment (BE) on the first- and last-mile travel modal choice. We select Singapore as a case study. The data incorporated for this work is extracted from the first- and last-mile trips to mass rapid transit (MRT) stations in the Household Interview Travel Survey of Singapore in 2012 with nearly 24 thousand samples. The BE indicators are quantified based on four "D" variables—Density, Diversity, Design, and Distance to transit. We also take into account sociodemographic and trip-specific variables. Mixed logit (ML) modelling frameworks are adopted to estimate the impact of BE and the heterogeneity of taste across the sample. Based on the availability of light rail transit (LRT) in different areas, two modeling structures are implemented with binary ML models for non-LRT areas where walk and bus are the available travel modes, and multinomial ML models for areas where LRT is an additional alternative. The modeling results shed light on the following findings: BE—especially the distance to MRT stations, transportation infrastructures, land-use mix and socioeconomic activities—significantly influences the first- and last-mile travel behaviors. For those who live or work close to MRT stations and in an area with high socioeconomic activities and land-use mix, they may have stronger preferences on walk for the first- and last-mile trips. The impact of physical BE (i.e. distance, infrastructures) is relatively homogeneous among the sample. While the impact of socioeconomic BE factors (i.e. floor space density, entropy) tend to vary across the sample.

Keywords: Built Environment (BE), Travel Behaviors, First and Last Mile, Mixed Logit Model

INTRODUCTION

Built environment (BE) is the man-made space in which people live, work, and recreate on a day-to-day basis (*1*). It encompasses urban design, land use, transportation system and patterns of human activity within the physical environment (*2*). BE can be quantified in several ways. One of the mostly widely used definition is the famously termed "D" variables by Ewing and Cervero (*3*). Past studies have revealed the impacts of urban form (*4–6*) and BE (*3*, *7*, *8*) on travel behavior, from which the findings provide profound reference for urban planning policy. Focusing on the access to and egress from transit facilities—so-called first- and last-mile trips, the studies on the influence of BE on first- and last-mile travel behaviors are, however, a few. Cervero et al. (*9*) found that people in denser places usually walk to transit stations. Similar conclusions are also drawn by Daniels and Mulley (*10*): walking distance to transit stops is mostly related to the mode of transit being accessed. Looking into the influence of street design and walkability on travel mode choice to transit stations, Park (*11*) yielded that better walkability increases the probability for transit users choosing walk instead of driving to the stations. The BE factors that these studies take into account are, however, not complete enough. A more comprehensive analysis on the relationship between BE and the first- and last-mile travel behavior can hardly be found. Tilahun et al. (*12*) conducted a wider range analysis of the last-mile issues in commuting trips incorporating the impact of BE. Nevertheless, by restricting to commuting trips, the findings may probably cause some bias in estimating the travel behavior. Traffic condition and demographic characteristics vary by country. The results of previous studies in America or Europe may not be suitable for Asian countries like Singapore. To fill the research gap, this study presents a comprehensive analysis on the impact of BE on first- and last-mile travel mode choice in Singapore, with four "D" characteristics—Density, Diversity, Design and Distance to transit (*3*)—to capture different perspectives of BE. In particular, the heterogeneity of the impact of BE, which is seldom considered in the literature, is also studied in this paper.

The first- and last-mile problem plays as an obstacle in promoting greater patronage of public transit. The distances between transit stations and the origins/destinations of passengers may sometimes be greater than the willingness to walk. One has to choose a feeder travel service to access to the transit station or may even use an alternative direct travel mode like personal vehicles, resulting in the systematic decrease of accessibility in urban areas. Some solutions have been proposed such as altering the location of transit stations to mixed-used activity centers; siting houses/workplaces near rail stations for improved proximity; constructing pedestrian footways, shaded corridors and bike lanes to improve walkability and connectivity (*9*, *12–16*). Most of these solutions tend to redesign or adjust the BE to improve the first- and last-mile experience.

In this study, we aim at investigating the impact of the BE on first- and last-mile modal choice. A mixed logit (ML) model framework is used to capture the heterogeneity of the impact of BE, which contributes to current literature in addition to previous studies (*9–12*). Singapore is selected as the case study area. Residents in Singapore heavily rely on public transport for daily travels. According to Household Interview Travel Survey (HITS) in 2012, during morning peak hours, 70% of commuters go to work by public transit, including Mass Rapid Transit (MRT) and bus. Thus, the first- and last-mile problem cannot be neglected in Singapore. In addition, the modal share of the first- and last-mile trips varies across the MRT stations (*17*), which may reflect the influence of BE in various locations. Such circumstances raise the importance in understanding the roles that BE plays on daily travel behaviors, especially in the context of Singapore.

The rest of this paper is organized as follows. The next section presents the processing and descriptive analysis of data. The methodology and model results are described in the third section. The final section discusses the findings and concludes the study.

DATA PROCESSING AND DESCRIPTIVE ANALYSIS

In the study, three categories of variables that may influence the first- and last-mile travel behaviors are collected. They are classified as sociodemographic variables (e.g., income, gender), trip-specific variables (e.g., travel time, travel cost), and BE variables. The data processing methods and descriptive analysis are illustrated in the following.

Household Interview Travel Survey

HITS is a paper-based household survey conducted every five years with a special focus on travel behavior in Singapore. The survey collects data on travel characteristics as well as individual sociodemographic information. The survey targets a sample size of at least 10,000 households, about 1% of the total number of households in Singapore (*18*). The sampled households are randomly selected by computer programs to ensure the representativeness of population. The data are collected through face-to-face interviews. The survey method follows the standard trip diary-based approach. In HITS, a trip is defined as a one-way journey completed for a specific purpose. On average 2.4 trips are collected for each respondent. The trip-specific characteristics of each travel segment (e.g., walking time to a bus stop, travel mode from home to the MRT station) are also recorded, which allows the identification of first- and last-mile trips.

In this study, a first/last mile trip is defined as the trip between an MRT station and the origin or destination of the journey. All travel records with MRT segments are first extracted from the HITS data. Then, the trip segments before and after the MRT trip are separated from the extracted records as the first- and last-mile samples. The sociodemographic information and trip-specific characteristics are collected as well. The samples with travel distance greater than 3 km are excluded, due to the fact that they are usually beyond the maximum service distance of a MRT station. These observations are not taken into account the first/last trips in this study. The exclusion of these data has little impact on the modal share of samples. In total 23,941 trips are extracted from the HITS database. The characteristics of travel segments are only recorded if a mode is actually used in the trip. We use Google Maps API to calculate the travel time and cost of the same trip by alternative modes, based on the departure time recorded in HITS.

Built Environment

The BE data are derived from the Singapore Land Authority (SLA) digitized cadastral dataset and the synthetic population data described in (*19*). The former contains detailed BE information, such as land use, postal codes and survey district numbers and boundaries. The latter are computed based on the iterative proportional fitting with two-stage approach (*19*), which contains the quantity and location of employment, residents and building floor space.

We divide Singapore into 1169 zones on the basis of Traffic Analysis Zones (MTZs) as shown in Figure 1. The average size of each MTZ is about 0.93 km^2. According to Ewing and Cervero (*7*), the BE impact is often studied in the neighborhood or activity center level in the literature. Thus, these divisions are reasonable for the BE variables calculation.

In this study, four "D" indicators are used to represent the BE: density, design, diversity, and distance to transit (*3*). Table 1 elaborates the measurement of each "D" variable. The variables are calculated in ArcGIS. Population density is defined as the total number of residents living in each MTZ divided by the corresponding residential area. The density of employment is estimated based on the methods proposed by Munshi (*20*), described as the total number of jobs in a MTZ divided by the economic activity area (e.g. commercial, industrial). The floor space density is calculated as the total building floor space in a specific MTZ divided by the area of this zone.

In terms of the diversity variables, we calculated the dissimilarity index and entropy based on Cervero and Kockelman (*21*). Six land-use categories are classified, including residential, commercial, institutional, industrial, recreational, and others (e.g. waterbody). We first latticed the island of Singapore into $100 \times 100\ m$ grid cells. The cells are set as the basic unit to calculate the diversity indices. Each grid cell is labeled with its most prevalent land use in terms of gross floor area for dissimilarity index calculation. A value of dissimilarity is then computed for each cell as the number of dissimilar land uses in labels of the eight "queen" neighborhood cells. The dissimilarity index of a MTZ is computed as the mean dissimilarity values of all internal cells. Greater dissimilarity index indicates higher level of land use mixture, typically considered as characteristics of smart growth *(22).* To calculate the entropy, a buffer of 800 m radius around the center of each grid cell was considered as the neighborhood area (*20*, *21*). The entropy of each grid cell is then estimated based on the land-use categories within the buffer. Similarly, the entropy of a MTZ is computed as the mean of the entropy value of all internal grid cells. The entropy index ranges from 0 to 1, where 1 signifies the perfect balance of land use with maximum heterogeneity and 0 indicates that there is only one land use in the neighborhood area (*21*).

The category of design is represented by the density of road length and road intersections, the kernel density of bus stops and MRT/LRT stations, and the ease of access index (EAI) to bus stops, MRT/LRT stations, and buildings. The density of road intersections represents the complexity of the road network and size of blocks. Expressways and walking paths are excluded from road length density calculation since they are seldom used by vehicles for first- and last-mile trips. The road length density is expected to have positive effect on the choice of motorized travel modes but to have negative effect on walking. Since high road length density means greater connectivity to motor vehicles (*23*). A kernel radius—2 km for bus and LRT stations and 3 km for MRT stations—and a distance decay function are assigned to each bus stop or MRT station for kernel density calculation (*22*). The value of the decay function reaches the peak at the source point and decreases smoothly as the distance increases within the kernel radius. The kernel value for each grid cell was computed as the sum of all overlapped function values. For each MTZ, the kernel density is calculated as the average kernel value for all grid cells in the zone. As opposed to point density, kernel density has the advantage of counting the BE effect not only on the zone containing the corresponding bus stops and MRT/LRT stations, but also on the neighborhood. Similar to the definition of Zhao (*23*), we define the EAI to represent the intensification of public transport services around an individual. It is the accumulation of the decay function value multiplied by footprint (i.e. the horizontal projection of the building) of the transit stops/stations within a 2 km distance threshold. The EAI to buildings is also defined by the same method, reflecting the potential of an individual to access surrounding socioeconomic activities. Since the EAI to MRT stations are expected to play an important role on first- and last-mile trips. Two decay functions are used to calculate the EAI to MRT stations (decayed by distance) and the walking-based EAI to MRT stations (decayed by walking time), respectively. The distance to transit stop is calculated as the direct distance between MRT station and the origin/destination. It is expected to be the most significant factor for people choosing between walking and motorized travel modes.

TABLE 1 Definition of BE Variables

Variable		Definition	Operationalization
Density	Population density	Population per residential square meter.	$D_p = P_r/A_r$, where D_p= population density, P_r= population in the residential area, A_r= area of residential land-use in a MTZ.
	Employment density	Employment per residential square meter.	$D_e = E_{ea}/A_{ea}$, where D_e= employment density, E_{ea}= total employment available in the economic activity area, A_r= area of economic activity land-use in a MTZ.
	Floor space density	Building floor space per square meter.	$D_{fs} = FS/A$, where D_{fs}= floor space density, FS= total floor space in a MTZ, A= area of the MTZ.
Diversity	Dissimilarity index	Proportion of dissimilar land use among grid cells within a MTZ (*21*).	Dissimilarity index = $I_d = \sum_{k=1}^{K}[\sum_{l=1}^{8}(X_l/8)]/K$, where K = the number of grid cells in a MTZ, $X_l = 1$ if the land use category of neighboring grid cell l differs from the grid cell k.
	Entropy	The mean entropy for land use categories among grid cells within an 800m radius of each grid cell within a MTZ (*21*).	Entropy = $E = \{\sum_{k=1}^{K}\left[\sum_j P_{jk}\ln(P_{jk})\right]/\ln(J)\}/K$, where K = the number of grid cells in a MTZ, J = the number of land-use classes. P_{jk} is the proportion of the area under the jth land-use type within the 800m radius surrounding grid cell k.
Design	Road length density	Road length per square meter.	Road length density = $D_r = L_{road}/A$, where L_{road}= road length, A= area of a MTZ.
	Road intersections density	The number of road intersections per square meter.	Road intersections density = $D_i = N_i/A$, where N_i= number of road intersections. A= area of a MTZ.
	Bus/MRT/LRT stations kernel density	Mean kernel density for Bus/MRT/LRT stations (*22*).	Kernel density = $KD = [\sum_{k=1}^{K}(\sum_m e^{-d_{km}})]/K$, where d_{km}= the distance (km) from grid cell k to Bus/MRT/LRT stations, K= the number of grid cells in a MTZ.
	EAI to MRT/LRT station	the ease of an individual to access surrounding MRT/LRT stations based on the gravity model. (*24*)	EAI to MRT/LRT station = $\sum_m A_{sm}e^{-d_{im}}$ where d_{im}= the distance (km) from individual i to MRT/LRT stations m within 2 km, A_{sm}= the footprint $(1000 \times \text{m}^2)$ of the MRT/LRT station m.
	EAI to bus stop	the ease of an individual to access surrounding bus stops based on the gravity model. (*24*)	EAI to bus stop = $\sum_m A_{bm}e^{-d_{im}}$ where d_{im}= the distance (km) from individual i to MRT/LRT stations m within 2 km, A_{bm}= the bus bay length (km) of the bus stop m.
	EAI to buildings	the ease of an individual to access surrounding buildings based on the gravity model. (*24*)	EAI to buildings = $\sum_m FS_{bm}e^{-d_{im}}$ where d_{im}= the distance (km) from individual i to building m within 2 km, FS_{bm}= the floor space $(1000 \times \text{m}^2)$ of the building m.
	Walking-based EAI to MRT station	the ease of an individual to access surrounding MRT stations by walking based on the gravity model. (*24*)	Walking-based EAI to MRT station = $\sum_m A_{sm}e^{-T_{im}}$ where T_{im}= the walking time $(1000 \times \text{sec.})$ of individual i to MRT station m within 2 km, A_{sm}= the footprint $(1000 \times \text{m}^2)$ of the MRT station m.
Distance to transit stop	Distance to MRT station	Euclidean distance between MRT station and the origin or destination	This is measure as Euclidean distance between MRT station and the origin or destination.

Descriptive Analysis

The origins and destinations of the sample of 23,491 trips are illustrated in Figure 1, which covered nearly all surrounding regions of the MRT lines, showing the spatial representativeness of the samples. Most of the green dots (choosing walk) are located near the MRT stations while the majority of yellow dots (choosing bus) are more distant from MRT stations, which hints that the distance to MRT stations play an important role in the first- and last-mile travel mode choice.

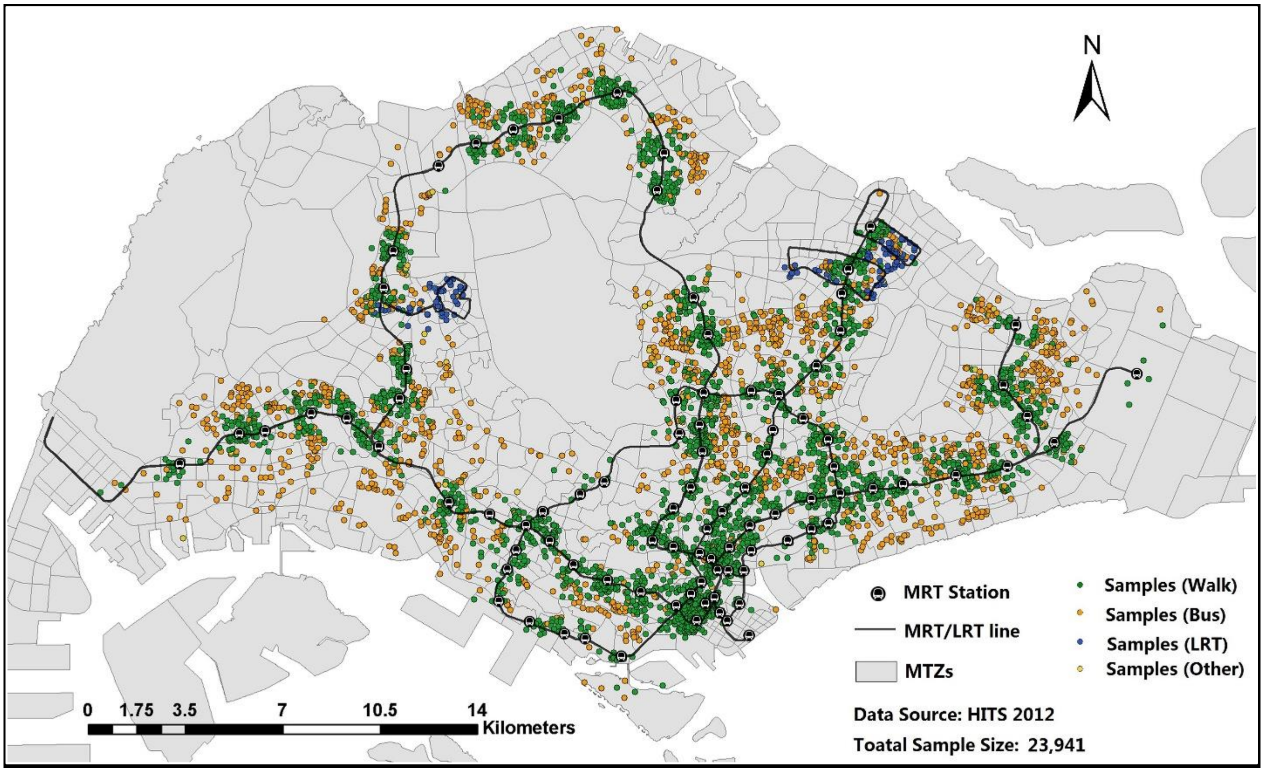

FIGURE 1 Samples Distribution in MTZs.

(Data source: authors' calculation)

In Singapore, walk and bus are the two major travel modes for the first- and last-mile trips. 69.06% of the respondents choose to walk to or from MRT stations, while the share of bus mode is 27.75%. The Light Rail Transit (LRT) mode accounts for 2.21% and the others only occupy 0.98%. The "other" modes (e.g. car driving and sharing) are not considered separately as the share is too small to estimate the variables impacts. Moreover, according to numerical test, discarding these modes has almost no influence on the modeling results. The small share of car may result by the fact that less than 10% of people in Singapore own private cars (*25*). If one chooses to drive, he/she is more likely to drive to the destination directly instead of driving to the MRT station for first/last mile only. Figure 2 presents the first- and last-mile modal share around different MRT stations. The modal share varies across MRT stations. The finding motivates us to explore the implied impact of various BE on the first- and last-mile travel behaviors. The LRT is only available around three specific MRT stations (Choa Chu Kang, Sengkang and Punggol). Therefore, two separated sample sets were created to model and to explore the travel behaviors in areas with and without LRT infrastructures, respectively.

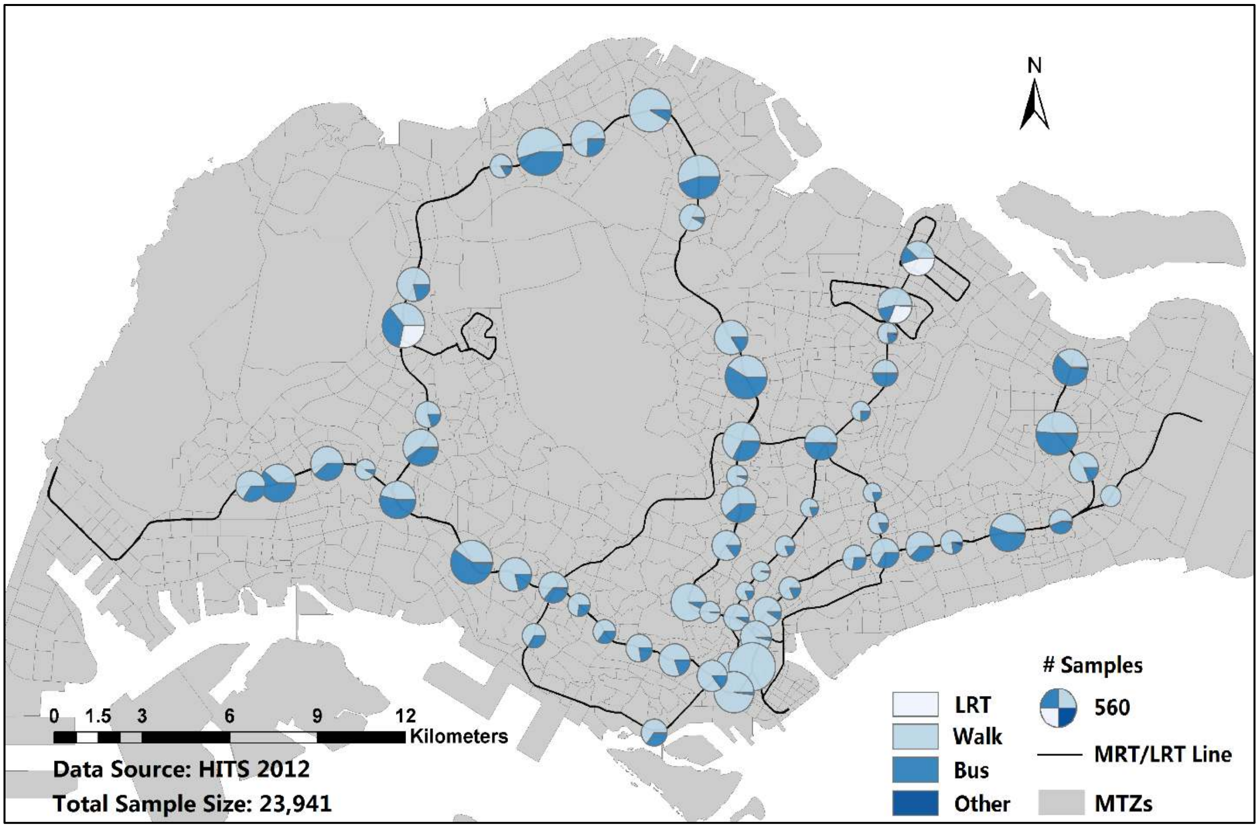

FIGURE 2 Modal Share of the First- and Last-mile Trips.
(Data source: authors' calculation)

Table 2 lists a summary of variables considered in this study. In terms of sociodemographic variables, the distributions of gender (48% male) and age (30% below 25; 44% 25–45 years old) in sample are reasonable. The BE variables of origin, destination and non-MRT station area of each first- and last-mile trip are calculated. For a first-mile trip, the origin refers to traveler's home or working place; and the destination means the corresponding MRT station. For a last-mile trip, the origin becomes the MRT station while the traveler's home or working place is the destination. In addition, the "non-MRT station area" represents the origin of first mile and/or the destination of last mile. The characteristics of socio-economic activity of an individual can be reflected in these areas. Since the first- and last-mile trips are mostly made within a short distance, the three categories of BE variables are highly correlated with each other. Therefore, only one of them is selected in the final model on the basis of goodness of fit and explanatory reasonability. The mean entropy of samples is relatively high (0.60), indicating higher land-use balance of areas near MRT stations. A wide variation in kernel density of transit stops can be observed, indicating the spatial differences in the construction of public transport facilities in Singapore. In terms of the trip-specific variables, we can find that one tends to choose motorized travel modes in relatively long distance trips but to choose to walk if the distance is shorter. The mean travel time of first- and last-mile trips is thus well controlled in a reasonable range (about 7–10 min).

TABLE 2 Summary of Variables

***(1)** Sociodemographic variables*

Variable name	Mean	Std. dev.

Live in public flat (Yes=1)	0.87	0.33
Singapore citizen (Yes=1)	0.83	0.38
Number of people in house	4.12	1.44
Have kids under six (Yes=1)	0.15	0.36
Age below 25 (Yes=1)	0.30	0.46
Age between 25 and 45 (Yes=1)	0.44	0.50
Gender (Male=1)	0.48	0.50
Have car license (Yes=1)	0.29	0.45
Full-time employment (Yes=1)	0.89	0.31
Income ($SG)	1842.00	1935.63
Commute trip (Yes=1)	0.43	0.50

(2) BE variables

Variable name	Mean	Std. dev.
*Non-MRT station area**		
Population density (people/m^2)	0.059	0.028
Employment density (jobs/m^2)	0.14	0.20
Floor space density (m^2/m^2)	2.51	3.73
EAI to bus stop	0.97	0.24
EAI to MRT station	17.49	21.62
EAI to building	6296.29	2899.25
EAI to LRT station	0.31	1.08
Walking-based EAI to MRT station	32.73	47.77
Bus stops kernel density	1.20	0..31
MRT stations kernel density	20.54	23.57
LRT stations kernel density	0.43	1.24
Entropy	0.60	0.15
Dissimilarity index	0.25	0.15
Road density (0.01m/m^2)	1.55	0.94
Road intersections density (Num./km^2)	17.51	18.37
Distance to MRT station (m)	713.35	623.66
Origin		
Population density (people/m^2)	0.065	0.34
Employment density (jobs/m^2)	0.50	0.20
Floor space density (m^2/m^2)	2.46	3.60
EAI to bus stop	1.00	0.25
EAI to LRT station	0.29	0.99
Bus stops kernel density	1.23	0.30
MRT stations kernel density	21.86	23.79
LRT stations kernel density	0.45	1.01
Entropy	0.61	0.15
Dissimilarity index	0.23	0.15
Road density (0.01m/m^2)	1.51	0.87
Road intersections density (Num./km^2)	19.39	18.87
Destination		
Population density (people/m^2)	0.064	0.33
Employment density (jobs/m^2)	0.50	0.20
Floor space density (m^2/m^2)	2.47	3.62
EAI to bus stop	0.96	0.26
EAI to LRT station	0.34	1.17
Bus stops kernel density	1.20	0.32

MRT stations kernel density	20.54	23.57
LRT stations kernel density	0.43	1.24
Entropy	0.60	0.15
Dissimilarity index	0.25	0.15
Road density ($0.01 m/m^2$)	1.51	0.87
Road intersections density (Num./km^2)	19.30	18.60

(3) Market share and Trip-specific variables

Model	Mode	Modal share (%)	Travel distance (m)		Travel time (min)	
			Mean	Std. dev.	Mean	Std. dev.
Binary ML	Walk	72.30	655.39	456.41	7.47	3.93
	Bus	26.74	2381.42	1229.09	10.34	5.14
	Other	0.96	*n.a.*	*n.a.*	*n.a.*	*n.a.*
Multinomial ML	Walk	52.74	763.16	432.05	6.97	3.92
	Bus	29.96	2045.41	985.53	8.87	4.19
	LRT	15.81	2642.75	1378.14	8.30	5.61
	Other	1.49	*n.a.*	*n.a.*	*n.a.*	*n.a.*

Note: n.a.: not applicable.

[*] *Some BE variables (e.g. EAI to MRT station) of non-MRT station area are not included in those of origin and destination since they are considered meaningless for first- and last-mile trips. For instance, the origin of a trip can be a MRT station. Thus, the EAI of an MRT station to another MRT station (i.e. EAI of origin to MRT station) is not meaningful of the first- or last-mile trips.*

MODEL RESULTS AND ANALYSIS

Methodologies

The choice behavior, as stated by Train (*26*), is based on the rational evaluation of all available alternatives and the maximization of utility. In this study, the alternatives consist of walk, bus and LRT. In addition to the BE variables, individual and household level factors are also assumed to affect the modal choice. Variables including gender, age, household size, income, travel purpose etc. are all tested to capture the influence of these variables on modal choice. The trip-specific variables (e.g. travel time and travel cost) are also taken into account in the models. In summary, we assume that the first- and last-mile modal choice is influenced by three different categories of factors: socio-demographics characteristics of respondents, BE at the origins, destinations and non-MRT station areas, and trip-specific variables.

The influence of BE on mode choice may vary across gender, household composition and socioeconomic groups (*27*, *28*). A household with an inclination towards a certain type of travel may self-select a residential location to use the preferred mode to travel (*28*). This phenomenon is commonly referred to as the residential self-selection problem. In this study, the impact of self-selection bias is assumed to be mitigated due to the Build-to-Order (BTO) policy in Singapore (*29*). People who want to buy a public flat developed by the Housing and Development Board need to apply and wait to be allocated. This policy results in the inflexible choice of residence, as well as the mixed sociodemographic characteristics in a residence, mitigating the self-selection bias.

The ML model is a highly flexible model that allows for estimating the random taste variation across individuals (*26*). The heterogeneity of the impact of BE is seldom estimated in past works. In this study, to estimate the mean impact and taste variation of BE, the ML modelling framework is adopted. Since the availability of LRT is constrained in three MRT station areas, individuals in these areas with EAI to LRT are selected to perform the multinomial ML model with three alternatives (i.e. walk, bus and LRT). Other samples are selected to perform a binary ML model (with walk and bus as alternatives). According to discrete choice theory, individuals are rational and choose the travel mode providing the highest utility. The utility function V_{ni} for person n choosing travel mode i is specified as follows (*26*, *30*):

$$V_{ni} = \alpha_i + \gamma_i' X_n + \theta_{ni}' B_n + \mu_{ni}' T_{ni} \tag{1}$$

where:

X_n is the vectors of sociodemographic characteristics of individual n;

B_n is the vectors of BE variables of individual n;

T_{ni} is the vectors of trip-specific attributes of mode i for individual n;

α_i is the alternative specific constant (ASC), capturing the inherent preference for mode i;

γ_i', θ_{ni}' and μ_{ni}' are the corresponding coefficients to be estimated.

According to the classic theory of ML modelling estimation (*26*, *30*, *31*), the probability of individual n choosing travel mode i can be expressed as:

$$P_{ni} = \int \frac{\exp(V_{ni})}{\sum_{k=1}^{K} \exp(V_{nk})} f(\beta)\, d\beta \tag{2}$$

where:

β is the vectors of coefficients which combines θ_{ni}' and μ_{ni}';

P_{ni} is the probability for individual n choosing mode i, simulated by taking draws of β and from the density $f(\beta)$, which is assumed to be normal distribution in this study;

K is the number of available alternatives for the individual.

Since we focus on the impact of BE and aim to improve the estimation efficiency of the proposed model, the taste variation of sociodemographic variables and ASCs are not measured in this study.

For analysis purpose, aggregate elasticity is often used to summarize the impact of an incremental change in a variable on the expected share of a group of decision makers (*30*). Derived from Train (*26*) and Ben-Akiva and Lerman (*30*), the percentage change in the expected probability for alternative i given a percentage change in the mth attribute of alternative j in population level is

$$E_{x_j^m}^{\bar{P}_i} = \frac{1}{N \cdot \bar{P}_i} \sum_{n=1}^{N} -x_{nj}^m \int \beta^m \frac{\exp(V_{ni})}{\sum_{k=1}^{K} \exp(V_{nk})} \frac{\exp(V_{nj})}{\sum_{k=1}^{K} \exp(V_{nk})} f(\beta)\, d\beta, \tag{3}$$

where:

N is the number of samples.

x_{nj}^m is the mth attribute of alternative j for person n.

β^m is the mth attribute element of β.

$\bar{P}_i$ is expected possibility of the group choosing alternative i, satisfying that

$$\bar{P}_i = \frac{\sum_{n=1}^{N} P_{ni}}{N}. \tag{4}$$

$E_{x_j^m}^{\bar{P}_i}$ is the aggregate elasticity of alternative i given the same increment of x_{nj}^m for each individual so that

$$\frac{\partial x_{nj}^m}{x_{nj}^m} = \frac{\partial x_{n'j}^m}{x_{n'j}^m} = \frac{\partial x_j^m}{x_j^m}, \qquad \text{for all } n, n' = 1,2,\dots,N. \tag{5}$$

where

$$x_j^m = \frac{\sum_{n=1}^{N} x_{nj}^m}{N}. \quad (6)$$

The models were estimated in pythonbiogeme with 1,000 random draws (*32*). The base category of the model is walk and all other modes are studied in comparison with it. Some variables described in the previous section were dropped from the final model due to the fact that they were found to be insignificant in explaining mode choice or were highly correlated with variables that were eventually included in the model. If two variables are highly correlated, the selection of variables is based on the goodness-of-fit and the sensitivity of policy, that is, we opted to keep variables that are explanatory and constructive for policy.

Binary ML model

After filtering out the samples with the "other" travel modes, in total 20,181 samples are used for the binary ML model estimation. Apart from the normal model, for comparative purposes, a control model without BE variables is tested. The estimation results of model with and without BE variables are shown in Table 3.1.

Both models in the table reveals high goodness-of-fit values, indicating the models are able to well predict the travel behavior. As shown in the table, by incorporating the BE variables, the goodness-of-fit value, the adjusted ρ^2, substantially improves from 0.733 to 0.832, which indicates the importance of BE impact in addition to the trip-specific and sociodemographic variables.

As in Table 3.1(a), the signs of all coefficients are consistent with the assumed effects. The ASC of bus (-9.510) is less than walk (fixed to 0), which suggests the negative preference on bus. Comparing with trip-specific and BE variables, sociodemographic variables have little effect on the first- and last-mile modal choice, which implies that the impact of self-selection could be minor. It is found that one with commuting purpose (e.g. work and education) is more likely to choose bus. This suggests that the previous research (*12*) which studied the BE impact on commuting trips may cause some bias. The model also suggests that travel time are important consideration in the choice of modes. Look into the estimates in Model (a). The absolute value of the coefficient of bus travel time is greater than that of walk, indicating that time spent by bus is more sensitive than the time spent by walk. The inconsistent values in Model (b) are possibly led by the fact that the parameter of bus travel time in Model (b) contains hidden positive effect of distance to MRT station. Thus, after adding the influence of distance in Model (a), the impact of bus travel time is adjusted. The standard deviation of bus travel time is statistically different from zero, which shows the various attitude towards bus travel time among the sample.

In terms of the impact of BE, the distance to MRT station plays an essential role in modal choice decision. The positive sign demonstrates higher probability of choosing bus with longer distance. By calculating the aggregate elasticity based on Eq. (3), if the distance to MRT station decreases by 10%, the share of walk on average tends to increase by 2.04% while the share of bus tends to decrease by 5.35 %. In addition, the density of building floor space in non-MRT areas is found to have negative effect on bus mode choice. This implies that the higher density of socioeconomic activities encourages people to walk. Meanwhile, the road length density has less sensitive impact on modal choice. If the road density increases by 10%, the share of walk then decreases by 0.07% while bus share increases by 0.17% on average. The only BE parameters found

TABLE 3 Modeling Estimates

***(1)** Estimation Results of Binary ML model*

Variable		(a) Model with BE variables			(b) Model without BE variables		
		Coefficient	t-test		Coefficient	t-test	
Walk							
Constant α	-	0	fixed		0	fixed	
Travel time	Mean	-0.567	-20.06	***	-0.580	-41.09	***
	†Std. Dev.	-0.115	0.11		0.383	0.23	
Bus							
Constant α	-	-9.510	-17.25	***	-6.38	-46.13	***
Travel time	Mean	-0.946	-16.09	***	-0.253	-21.73	***
	Std. Dev.	0.250	12.42	***	0.058	4.97	***
Commute trip (Yes=1)	-	0.235	1.81	*	0.243	3.81	***
Distance to MRT station	†Mean	1.160	15.95	***	-	-	
	‡Std. Dev.	0.102	0.04		-	-	
EAI to Bus stop (Origin)	Mean	2.650	6.98	***	-	-	
	Std. Dev.	0.037	0.28		-	-	
Floor space density (Non-MRT station area)	Mean	-0.329	-4.53	***	-	-	
	Std. Dev.	0.146	3.23	***	-	-	
Walking-based EAI to MRT station	Mean	-0.039	-6.43	***	-	-	
	Std. Dev.	0.027	6.81	***	-	-	
Road density (Non-MRT station area)	Mean	0.144	1.75	*	-	-	
	Std. Dev.	0.362	0.06		-	-	
Statistics							
Observations		20181			20181		
Rho squared		0.832			0.736		
Adjusted Rho squared		0.831			0.735		

*Note:† Data scaled by 100, ‡Data scaled by 10,000; ***: Significant at 99% level; **: Significant at 95% level, *: Significant at 90% level*

(2) *Estimation Results of Multinomial ML model*

Variable		(a) Model with BE variables			(b) Model without BE variables		
		Coefficient	t-test		Coefficient	t-test	
Walk							
Constant α	-	0	fixed		0	fixed	
Travel time	Mean	-0.835	-4.06	***	-1.260	-6.42	***
	[†]Std. Dev.	0.144	2.77	**	0.235	4.28	***
Bus							
Constant α	-	-3.860	-1.46		-7.290	-6.92	***
Travel time	Mean	-1.850	-3.68	***	-0.904	-5.57	***
	Std. Dev.	0.392	2.99	***	0.154	2.70	**
Distance to MRT station	[†]Mean	2.450	3.58	***	-	-	
	[‡]Std. Dev.	1.430	0.16		-	-	
Entropy (Non-MRT station area)	Mean	-15.40	-2.95	***	-	-	
	Std. Dev.	0.439	0.34		-	-	
EAI to bus stops (Origin)	Mean	3.020	2.76	**	-	-	
	Std. Dev.	0.141	0.19		-	-	
LRT							
Constant α	-	11.90	1.43		-7.790	-6.35	***
Travel time	Mean	-3.230	-2.71	**	-1.130	-6.11	***
	Std. Dev.	0.540	2.29	**	0.008	0.11	
Distance to MRT station	[†]Mean	3.250	2.69	**	-	-	
	[†]Std. Dev.	0.032	0.23		-	-	
Entropy (Non-MRT station area)	Mean	-44.40	-2.38	**	-	-	
	Std. Dev.	3.600	1.73	*	-	-	
Statistics							
Observations		2373			2373		
Rho squared		0.891			0.816		
Adjusted Rho squared		0.885			0.813		

*Note: [†] Data scaled by 100, [‡]Data scaled by 10,000; *** Significant at 99% level; **: Significant at 95% level, *: Significant at 90% level* to vary across the individuals is walking-based EAI to MRT station and floor space density. The former indicates the variation of tastes on walking in the sample. The latter suggests the various attractiveness of socioeconomic activities. These two variables are both related to human activities, we may find that the impact of socioeconomic-related BE factors tend to vary across the sample. However, the insignificant standard deviation of other parameters (e.g. distance to MRT stations, road length density) indicates more homogeneous impact of physical BE on the sample.

Multinomial ML model

To model the impact of BE in the areas with LRT, the multinomial ML models are adopted with 2,373 sample trip segments in 84 MTZs. The estimation results are listed in Table 3.2(a). Similar to the binary ML model, a high goodness of fit value (adjusted $\rho^2 = 0.885$) is obtained in the multinomial ML model considering BE effect, which shows these variables can well describe people's behavior. A control model without BE variables is estimated for comparative purpose. The estimates are shown in Table 3.2(b). A substantial decrease of goodness of fit (adjusted ρ^2 decrease from 0.885 to 0.813) can be found after discarding BE variables, which emphasizes the importance of BE factors on people's modal choice.

Look into the estimates in Table 3.2(a). The coefficients of bus travel time, distance to MRT stations, and EAI to bus stops provide the same implication as that in Table 3.1(a), indicating the robust effect of these variables on modal choice behavior. The ASCs of bus and LRT modes are not statistically significant, which means people have no inherent preference for these three modes when LRT is available.

As for LRT mode, the positive signs of distance to MRT station shows that, similar to bus, one tends to use LRT when they are distant from the MRT station. The distance coefficients of LRT (3.250) and bus (2.450) suggest that, keeping all other variables constant, the increases distance from MRT station may encourage more people to use LRT than bus in the areas where the LRT is available. It is possible due to the fact that LRT can directly take the passengers to the MRT station, avoiding additional walking for interchange and the potential encounter of traffic congestions. According to the aggregate elasticity, if the distance to MRT station increases by 10% for each individual, the share of walk on average may decrease by 4.64% while bus and LRT may increase by 6.45% and 12.15%, respectively. Besides, the entropy is found to have negative effect on the utility of bus and LRT modes, which suggests that the high land use mix will encourage more people to walk. This finding corresponds to the previous study (*14*).

Similar to the binary ML model, the standard deviations of physical BE variables are not significant in this model, while the coefficient of entropy for LRT mode is found to vary across the sample. The finding is consistent with the heterogeneity of impact of physical and socioeconomic BE in binary ML model.

CONCLUSION AND DISCUSSION

The paper studies the impacts of BE on first- and last-mile travel mode choice based on the discrete choice model. We select Singapore as a case study. 23,941 observations of first- and last-mile trips were extracted from the HITS database. The BE factors were quantified using the four "D" variables proposed by Ewing and Cervero (*3*). In addition, sociodemographic variables and trip-specific variables were also taken into account in this work. To estimate the impact of BE and variation of taste, the ML modelling frameworks are adopted. Since the availability of LRT may cause a significant influence on travel behavior, two separated sample sets were used for performing a binary ML model (with walk and bus modes) and a multinomial ML model (with walk, bus and LRT modes), respectively. The models reveal following findings. 1) BE—especially the distance to MRT stations, transportation infrastructures, land use mix and socioeconomic activities—significantly influences the first- and last-mile travel behaviors. 2) For those who live or work close to MRT stations and in an area with high socioeconomic activities and land use mix, they may have stronger preferences toward walk for the first- and last-mile trips. 3) The impact of physical BE (i.e. distance, infrastructures) is relatively homogeneous across the sample. While the impact of socioeconomic-related BE (i.e. floor space density, entropy) varies.

There are several policy implications associated with the modeling results. From the city design point of view, the increase of the probabilities of choosing walking would come from design and building more compact communities with higher building floor space density and closer to the MRT stations. Recent studies have revealed that walking can reduce risk of death and the burden of important chronic conditions (*33–35*). This research provides meaningful suggestions for improving public health from the BE angle by promoting walking modes. Besides, the results also offer some suggestions for planning authorities to balance the demand of bus. For those who live away from MRT stations, a bus system with high density of bus stops, better accessibility to the stops, and higher road network density should be provided to meet their first- and last-mile travel demands. In addition, the results also offer a reference for the prospective implementation of autonomous vehicles (AV). The Land Transport Authority (LTA) of Singapore released the Singapore Autonomous Vehicle Initiative (SAVI) to explore the technology, application, and solutions with AV in Singapore (*36*). Since AVs and buses are both motorized and shared travel modes for first- and last-mile trips, the areas with high first/last mile travel demand by bus may also imply high potential demand of AVs in the future. Therefore, the model results offer some suggestions for AV deployment and infrastructures installation with consideration of BE to balance the use of different modes.

This study can be further improved from the following aspects. The first one pertains to assumption of ignoring the self-selection bias. Due to the BTO policy, the common method with sociodemographic variables as the control may not be applicable here. Therefore, this assumption can only be further tested with more attitudinal data and more advanced modeling approach, which beyond the scope of the present study. Another path to improve of this work relates to coping with multi-collinearity of data. The multi-collinearity problem results in the discard of several variables such as the MRT station density. Future work can be done by applying dimension reduction method (e.g. factor analysis, principal component analysis) to extract latent variables to better illustrate the impact of BE.

ACKNOWLEDGEMENT

The research is supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under CREATE programme, Singapore-MIT Alliance for Research and Technology (SMART) Centre, Future Urban Mobility (FM) Interdisciplinary Research Group. The authors thank Anson F. Stewart for his insightful comments and proofreading, and thank Daya Zhang for editing the figures. The first author also thanks the TOP OPEN program and Initiative Scientific Research Program of Tsinghua University, China, for financial support (20161080166).

REFERENCES

1. Roof, K. Public Health: Seattle and King County's Push for the BE. *Journal of environmental health*, Vol. 71, No. 1, 2008, p. 24.

2. Handy, S. L., M. G. Boarnet, R. Ewing, and R. E. Killingsworth. How the BE Affects Physical Activity: Views from Urban Planning. *American journal of preventive medicine*, Vol. 23, No. 2, 2002, pp. 64–73.
3. Ewing, R., and R. Cervero. Travel and the BE: A Meta-Analysis. *Journal of the American planning association*, Vol. 76, No. 3, 2010, pp. 265–294.
4. Stead, D., and S. Marshall. The Relationships between Urban Form and Travel Patterns. An International Review and Evaluation. *European Journal of Transport and Infrastructure Research*, Vol. 1, No. 2, 2001, pp. 113–141.
5. Dieleman, F. M., M. Dijst, and G. Burghouwt. Urban Form and Travel Behaviour: Micro-Level Household Attributes and Residential Context. *Urban Studies*, Vol. 39, No. 3, 2002, pp. 507–527.
6. Frank, L., M. Bradley, S. Kavage, J. Chapman, and T. K. Lawton. Urban Form, Travel Time, and Cost Relationships with Tour Complexity and Mode Choice. *Transportation*, Vol. 35, No. 1, 2008, pp. 37–54.
7. Ewing, R., and R. Cervero. Travel and the BE: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1780, 2001, pp. 87–114.
8. Cao, X. J., P. L. Mokhtarian, and S. L. Handy. The Relationship between the BE and Nonwork Travel: A Case Study of Northern California. *Transportation Research Part A: Policy and Practice*, Vol. 43, No. 5, 2009, pp. 548–559.
9. Cervero, R., A. Round, T. Goldman, and K.-L. Wu. Rail Access Modes and Catchment Areas for the BART System. *University of California Transportation Center*, 1995.
10. Daniels, R., and C. Mulley. Explaining Walking Distance to Public Transport: The Dominance of Public Transport Supply. *Journal of Transport and Land Use*, Vol. 6, No. 2, 2013, pp. 5–20.
11. Park, S. *Defining, Measuring, and Evaluating Path Walkability, and Testing Its Impacts on Transit Users' Mode Choice and Walking Distance to the Station*. University of California, Berkeley, 2008.
12. Tilahun, N., P. V. Thakuriah, M. Li, and Y. Keita. Transit Use and the Work Commute: Analyzing the Role of Last Mile Issues. *Journal of Transport Geography*, Vol. 54, 2016, pp. 359–368.
13. Naharudin, N., M. S. S. Ahamad, and A. F. M. Sadullah. Assessing Criteria of the Pedestrian-Friendly First/Last Mile Transit Journey by Using Analytical Network Process (ANP) Group Judgment. *PROCEEDINGS ICOSH-UKM 2017*, p. 203.
14. Ryan, S., and L. F. Frank. Pedestrian Environments and Transit Ridership. *Journal of Public Transportation*, Vol. 12, No. 1, 2009, p. 3.
15. Loutzenheiser, D. Pedestrian Access to Transit: Model of Walk Trips and Their Design and Urban Form Determinants around Bay Area Rapid Transit Stations. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1604, 1997, pp. 40–49.
16. Jiang, Y., P. C. Zegras, and S. Mehndiratta. Walk the Line: Station Context, Corridor Type and Bus Rapid Transit Walk Access in Jinan, China. *Journal of Transport Geography*, Vol. 20, No. 1, 2012, pp. 1–14.
17. Shen, Y., H. Zhang, and J. Zhao. *Embedding Autonomous Vehicle Sharing in Public Transit System: An Example of Last-Mile Problem*. Transportation Research Board,Washington, D.C., 2017.
18. Singapore Land Transport Authority. *2012 Household Interview Travel Survey Background Information*. 2012.
19. Zhu, Y., and J. Ferreira. Synthetic Population Generation at Disaggregated Spatial Scales for Land Use and Transportation Microsimulation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2429, 2014, pp. 168–177.
20. Munshi, T. BE and Mode Choice Relationship for Commute Travel in the City of Rajkot, India. *Transportation Research Part D: Transport and Environment*, Vol. 44, 2016, pp. 239–253.
21. Cervero, R., and K. Kockelman. Travel Demand and the 3Ds: Density, Diversity, and Design. *Transportation Research Part D: Transport and Environment*, Vol. 2, No. 3, 1997, pp. 199–219.
22. Tracy, A. J., P. Su, A. W. Sadek, and Q. Wang. Assessing the Impact of the BE on Travel Behavior: A Case Study of Buffalo, New York. *Transportation*, Vol. 38, No. 4, 2011, pp. 663–678.

23. Zhao, P. The Impact of the BE on Individual Workers' Commuting Behavior in Beijing. *International Journal of Sustainable Transportation*, Vol. 7, No. 5, 2013, pp. 389–415.
24. Minocha, I., P. Sriraj, P. Metaxatos, and P. Thakuriah. Analysis of Transit Quality of Service and Employment Accessibility for the Greater Chicago, Illinois, Region. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2042, 2008, pp. 20–29.
25. Singapore Land Transport Authority. *Singapore Land Transport: Statistics In Brief 2012*. 2012.
26. Train, K. E. *Discrete Choice Methods with Simulation*. Cambridge university press, 2009.
27. Badoe, D. A., and E. J. Miller. Transportation–land-Use Interaction: Empirical Findings in North America, and Their Implications for Modeling. *Transportation Research Part D: Transport and Environment*, Vol. 5, No. 4, 2000, pp. 235–263.
28. Schwanen, T., and P. L. Mokhtarian. What Affects Commute Mode Choice: Neighborhood Physical Structure or Preferences toward Neighborhoods? *Journal of Transport Geography*, Vol. 13, No. 1, 2005, pp. 83–99. https://doi.org/http://dx.doi.org/10.1016/j.jtrangeo.2004.11.001.
29. Housing & Development Board. *Build-To-Order (BTO)*. http://www.hdb.gov.sg/cs/infoweb/residential/buying-a-flat/new/bto-sbf. Accessed Jul. 30, 2017.
30. Ben-Akiva, M. E., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT press, 1985.
31. Hensher, D. A., J. M. Rose, and W. H. Greene. *Applied Choice Analysis: A Primer*. Cambridge University Press, 2005.
32. Bierlaire, M. *PythonBiogeme: A Short Introduction*. 2016.
33. Pucher, J., R. Buehler, D. R. Bassett, and A. L. Dannenberg. Walking and Cycling to Health: A Comparative Analysis of City, State, and International Data. *American journal of public health*, Vol. 100, No. 10, 2010, pp. 1986–1992.
34. Hamer, M., and Y. Chida. Active Commuting and Cardiovascular Risk: A Meta-Analytic Review. *Preventive medicine*, Vol. 46, No. 1, 2008, pp. 9–13.
35. Celis-Morales, C. A., D. M. Lyall, P. Welsh, J. Anderson, L. Steell, Y. Guo, R. Maldonado, D. F. Mackay, J. P. Pell, and N. Sattar. Association between Active Commuting and Incident Cardiovascular Disease, Cancer, and Mortality: Prospective Cohort Study. *BMJ*, Vol. 357, 2017, p. j1456.
36. Singapore Land Transport Authority. *Singapore Autonomous Vehicle Initiative*. https://www.lta.gov.sg/content/ltaweb/en/roads-and-motoring/managing-traffic-and-congestion/intelligent-transport-systems/savi.html. Accessed Jul. 25, 2017.