Transportation Research Part D 104 (2022) 103197

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

# Quantifying residential self-selection effects on commuting mode choice: A natural experiment

Yiling Deng [a], Pengjun Zhao [b,c,*]

[a] School of Design and Architecture, Zhejiang University of Technology, Hangzhou 310023, China
[b] School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen 518055, China
[c] College of Urban and Environmental Sciences, Peking University, Beijing 100871, China

ARTICLE INFO

*Keywords:*
Residential self-selection
Built environment
Commuting mode choice
Natural experiment
Propensity score weighting
Simultaneous equation model

ABSTRACT

Though the impact of residential self-selection on travel behavior has been widely discussed in transport studies, few studies have examined its effect on commuting mode choice, particularly when considering the mediating effects of commuting distance and car ownership. By using survey data from 19 large cities in China and taking advantage of the largely exogenous residential locations of social housing residents, we employed a natural experiment framework that combines propensity score weighting and simultaneous equation models to investigate the effects of the built environment and residential self-selection on commuting mode choice. The results show that most direct and indirect effects through commuting distance and car ownership are significant. The total effects of the built environment and residential self-selection have the same coefficient signs for car and regular bike modes, while they have opposite coefficient signs for public transit and e-bike modes.

## 1. Introduction

The impact of the built environment on travel behavior has received considerable scholarly interest, given that it can provide important guidance for policy-making and planning practice. Residential self-selection is one of the most important theoretical mechanisms to explain the link between the built environment and travel behavior. Residential self-selection refers to people' tendency to choose their residential environments based on their travel preferences (van Wee, 2009). Ignoring residential self-selection may bias the estimates of the impact of the built environment on travel behavior, which in most cases is overestimated but may also be underestimated (Cao et al., 2009).

There is extensive literature on residential self-selection (Guan et al., 2020). On the one hand, studies have examined possible complex associations among the built environment, travel preferences, and travel behavior (Lin et al., 2017). In addition to confirming the existence of residential self-selection, recent research has found that people may adapt their travel preferences based on their residential environments (also termed "residential determination" in some studies) (Lin et al., 2017; Ewing et al., 2016). On the other hand, studies have compared the magnitude and direction of the built environment and residential self-selection. Some research has further quantified the proportions of the effects of the built environment and residential self-selection on travel behavior (also known as the built environment proportion) (van Herick and Mokhtarian, 2020). Although existing studies have estimated the effect of residential self-selection on various outcome variables (Cheng et al., 2019; Yu et al., 2019; Deng and Yan, 2019; Liu et al., 2018; Lee

* Corresponding author at: School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen 518055, China.
*E-mail address:* pengjun.zhao@pku.edu.cn (P. Zhao).

https://doi.org/10.1016/j.trd.2022.103197
Received 9 August 2021; Received in revised form 24 January 2022; Accepted 24 January 2022

and Guhathakurta, 2018), few studies have examined its effects on commuting mode choice, particularly when considering the indirect effects through commuting distance and car ownership. Yet understanding the effect of residential self-selection on commuting mode choice is critical for urban and transportation policy-making and planning practice.

Previous methodologies for estimating the effect of residential self-selection, whether the traditional statistical control method (Cao and Ettema, 2014; Cao, 2015; Ettema and Nieuwenhuis, 2017) or the more advanced propensity score method (Yang et al., 2017; Cheng et al., 2019; Deng and Yan, 2019), have relied on a set of attitudinal variables. The problem faced by these studies is that a complete capture of attitudinal variables is impossible, and any omitted attitudinal variables may bias the estimates of the effects of the built environment and residential self-selection (Bohte et al., 2009). A natural experiment is another way to distinguish the effects of the built environment and residential self-selection. In many countries, especially in developing countries, the residential choice is highly constrained because many properties are owned by government authorities or work units (Guan and Wang, 2020). For example, affordable housing is assigned by the government based on eligibility and household size, and is usually located in remote suburban areas. Work-unit housing is assigned by the work unit based on job rank and marital status, and is usually located adjacent to workplaces. These social housing residents have little flexibility to choose residential locations based on their travel preferences compared to those living in private housing. By applying the natural experiment method, we can calculate the proportion of residential self-selection by comparing the travel behavior of private housing residents with those of social housing residents having similar sociodemographic attributes. It is not necessary to use any attitudinal variables in this process.

In this study, we adopted a natural experiment to divide residents into four groups (i.e., suburban private housing, suburban social housing, urban private housing, and urban social housing) based on whether they lived in the urban or suburban neighborhoods and whether they had the flexibility to choose where to live. We used propensity score weighting and simultaneous equation models to control for sociodemographic differences among the four groups, as well as to estimate the effects of the built environment and residential self-selection on commuting mode choice. The contributions of this paper lie in the following distinguishing features: 1) the use of simultaneous equation models to reveal the direct, indirect, and total effects of the built environment and residential self-selection on commuting mode choice; 2) the use of propensity score weighting and simultaneous equation models to obtain doubly robust estimates; 3) the possibility to offer more generalizable results by using data from 19 large cities across the whole of China.

The rest of the paper is structured as follows: Section 2 provides a literature review to establish the background for the study; Section 3 introduces the methodology; Section 4 presents the data and variables; Section 5 presents the analysis of the results; Section 6 discusses the empirical findings and justifies the methodology and results; the last section draws conclusions.

## 2. Literature review

Numerous studies have explored the relationships between the built environment, travel attitudes, and travel behavior (Guan et al., 2020). While the impacts of the built environment and travel attitudes on travel behavior have long been acknowledged, the relationship between the built environment and travel attitudes is controversial. Most studies have concluded that travel attitudes may influence the built environment, that is, people self-select their residential built environment according to their travel preferences. Some recent studies have argued that the built environment may influence travel attitudes (sometimes termed "residential determination") because people may adapt their travel preferences based on their residential built environment. It is difficult to distinguish which of these two effects is true or whether both exist because travel attitudes at the time of residential choice are difficult to observe (Lin et al., 2017; Ewing et al., 2016).

The existence of residential self-selection may bias the causal inference of the effect of the built environment on travel behavior. Various approaches have been employed to control for residential self-selection, with the most frequently applied being statistical control (Cao and Ettema, 2014; Cao, 2015; Ettema and Nieuwenhuis, 2017), propensity score (Yang et al., 2017; Cheng et al., 2019; Deng and Yan, 2019), and sample selection approaches (Cao, 2009; Bhat and Eluru, 2009; Zhang et al., 2017). The statistical control approach directly incorporates attitudinal variables as explanatory variables to eliminate bias due to the correlation of built environment variables with attitudinal variables. The propensity score approach models the choice between neighborhood types and uses the predicted choice probabilities as propensity scores to match cases to correct for selection bias. The sample selection approach models the choice between neighborhood types and incorporates the inverse Mills ratio from the selection model into the regression model to correct for selection bias. One aspect that these studies have in common is the need for explicit attitudinal control variables to treat the effect of residential self-selection. The reliability of these methods depends heavily on whether the attitudinal variables are fully controlled. Previous studies have used different attitudinal control variables depending on the research purpose or the data availability. Commonly used attitudinal control variables include attitudes toward travel in general, specific transportation modes, travel-related policies, and residential built environments (Guan et al., 2020). Note that these attitudes do not always affect the built environment and travel behavior. Travel preferences are often not readily translated into residential choices, and only residential preferences that can manifest themselves as travel preferences affect travel behavior (Wolday et al., 2019). The problem faced by these studies is that a complete capture of attitudinal variables is difficult, and omitted attitudinal variables may bias the estimates of the effects of the built environment and residential self-selection.

A natural experiment is another way to distinguish the effect of residential self-selection (Lin et al., 2017). Since social housing residents have little flexibility in choosing their residential built environments and are usually unable to realize their preferences, exploring the relationship between the built environment and travel behavior of social housing residents can rule out the bias of residential self-selection (Wang and Lin, 2014). By using a natural experiment to take advantage of the largely exogenous residential locations of people living in social housing, Zang et al. (2019) distinguished the effect of residential self-selection from the effect of the built environment without using any attitudinal variables. However, the application of a natural experiment to estimate the effect of residential self-selection has been limited.

Existing studies have estimated the effect of residential self-selection on various outcome variables, including travel behavior in general such as travel frequency (Deng and Yan, 2019), travel duration (Cheng et al., 2019), and trip chaining behavior (Liu et al., 2018); mode-specific travel behavior such as active travel (Haybatollahi et al., 2015), transit use (Yu et al., 2019), and vehicle use (Lee and Guhathakurta, 2018); as well as other issues such as residential choice (Humphreys and Ahern, 2019), car ownership (Liu et al., 2018), and transport emissions (Cao and Yang, 2017). However, few studies have focused on the effect of residential self-selection on commuting mode choice. Moreover, some studies that investigated the influence of various factors (e.g., sociodemographics, travel attitudes, built environment, and social context) on commuting mode choice have ignored the impact of residential self-selection (Clark et al., 2016; Sun et al., 2017; Ababio-Donkor et al., 2020; Bautista-Hernández, 2021). From an urban and transportation policy perspective, focusing on commuting mode choice is important because changing commuting modes can reduce the negative impacts of commuting such as traffic congestion and air pollution.

Residential self-selection has been examined mainly in developed areas such as the United States and Europe (Guan et al., 2020). Some lessons are not directly transferable to China due to its unique characteristics (Wang and Lin, 2014). In recent years, there have been increasing studies on residential self-selection in China (Guan and Wang, 2019; Cheng et al., 2019; Zhang and Zhang, 2018; Zhao and Zhang, 2018; Cao and Yang, 2017). These studies have provided insights into the mechanisms and impacts of residential self-selection. For example, they have confirmed the existence of residential self-selection, but often on a smaller scale compared to the built environment; they have also suggested the importance of housing source and specific residential choice preferences (e.g., short commutes and proximity to markets for daily goods shopping) in residential self-selection research. Yet, despite this growing body of research, most of these studies used data from a single city. Therefore, the results may exclusively reflect a single case study and cannot be generalized or provide policy implications for other cities. Using multi-city data to investigate whether the results are consistent with our hypothesized causal mechanisms across a wide range of cities may yield more generalizable impact results than case studies of one or a few cities.

## 3. Methodology

The methodological framework of this study is shown in Fig. 1. The unique housing system in China dictates that housing source has an important implication for the flexibility of residential choice, with only private housing residents able to realize their residential choice needs, while social housing residents cannot do so (Wang and Lin, 2014). We first divided the residents into four groups (i.e., suburban private housing, suburban social housing, urban private housing, and urban social housing), according to the residential location and housing source. These four groups of residents constitute a natural experiment to discern the effects of the built environment and residential self-selection.

We then applied propensity score weighting (PSW) to eliminate the sociodemographic differences among the four groups. The propensity score is the probability that a person is assigned to a treatment group rather than a control group given a set of covariates. Weighting the treatment and control groups using propensity scores eliminates heterogeneity between groups, that is, the covariates become similar across groups, thus making the groups comparable. In this study, we employed the generalized boosted machine rather than the traditional multinomial logit model to obtain propensity scores. Evidence suggests that the generalized boosted machine outperforms the multinomial logit model in terms of balance because the generalized boosted machine can automatically handle covariate polynomial terms and interaction terms and capture nonlinear relationships between group assignments and covariates (Lee et al., 2009). Propensity score weighting was implemented by using the twang package in R. The effects we estimated are the average treatment effects on the population (ATE), which is the average expected difference between the outcomes of the population assigned to one group versus another.

The literature suggests that reciprocal influences may exist between the built environment and travel preferences (Lin et al., 2017; Ewing et al., 2016). In this study, we only used the residential self-selection framework, i.e., travel preferences affecting the built environment, for the following two reasons. First, Guan et al. (2020) suggested that explicit identification requires three-wave panel household surveys. With only cross-sectional data, we cannot explicitly identify which of the residential self-selection framework,

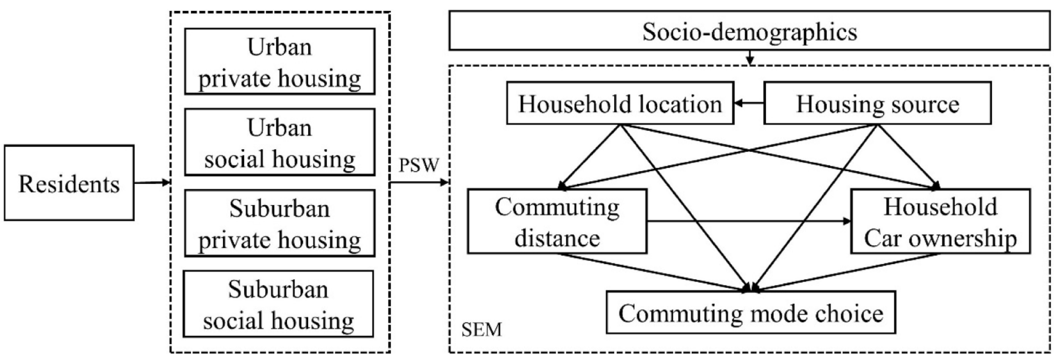


**Fig. 1.** Methodological framework.

residential determination framework, or causal reciprocal framework is more plausible. Second, residential self-selection is represented by housing source in this study, which is a dummy variable that indicates whether a resident is in private and social housing; the residential determination framework cannot be estimated because both private and social housing residents could have a residential determination effect, or in other words, this effect exists regardless of whether the dummy variable is 0 or 1.

Commuting mode choice is influenced by socio-demographics, travel characteristics (e.g., commuting distance and car ownership), the built environment, and residential self-selection (Ding et al., 2017), while commuting distance and car ownership themselves may also be influenced by socio-demographics, the built environment, and residential self-selection (Van Acker and Witlox, 2011). Specifically, commuting distance and car ownership may serve as mediating variables between the effects of socio-demographics, the built environment, and residential self-selection on commuting mode choice. We used single-directional relationships, i.e., people first choose household location and workplace (commuting distance is then determined), then choose whether to own a car, and finally choose the commuting mode. The logic of sequential decision-making from long-term to medium-term to short-term is widely applied in travel demand models. Although the reciprocal influences may exist between these endogenous variables, such as commuting distance influencing household locations, we believe the influence is limited. Due to the traditional preference for property ownership and the high investment value of properties, Chinese residents usually tend to prioritize the purchase of properties over cars. In addition, considering the reciprocal influences causes statistical identification issues and makes the models very complex and deviates from the focus of our study. We used simultaneous equation models (SEM) to estimate the direct and indirect relationships among these variables. Each simultaneous equation model consists of a set of regression equations that follow the path diagram shown in Fig. 1.

In this study, commuting mode choice was represented by five binary commuting mode variables and therefore five simultaneous equation models were estimated correspondingly. Mode choice is frequently analyzed with the multinomial logit (MNL) model. However, the multinomial logit model is an exponential model, and the coefficients represent changes in the log odds of the outcome, which is not straightforward and cannot be compared with other studies that use difference-in-means type analysis. In contrast, the simultaneous equation model is a set of linear models, and the coefficients represent changes in the outcome, which are more easily interpreted and compared. Note that the simultaneous equation model has a disadvantage that the predicted values of the binary commuting modes may fall outside the range of 0 and 1. However, we aimed to assess the effects of the built environment and residential self-selection rather than to predict the probability of individuals' mode choice, so this disadvantage was ignored.

We used diagonally weighted least squares instead of the commonly used maximum likelihood to estimate the simultaneous

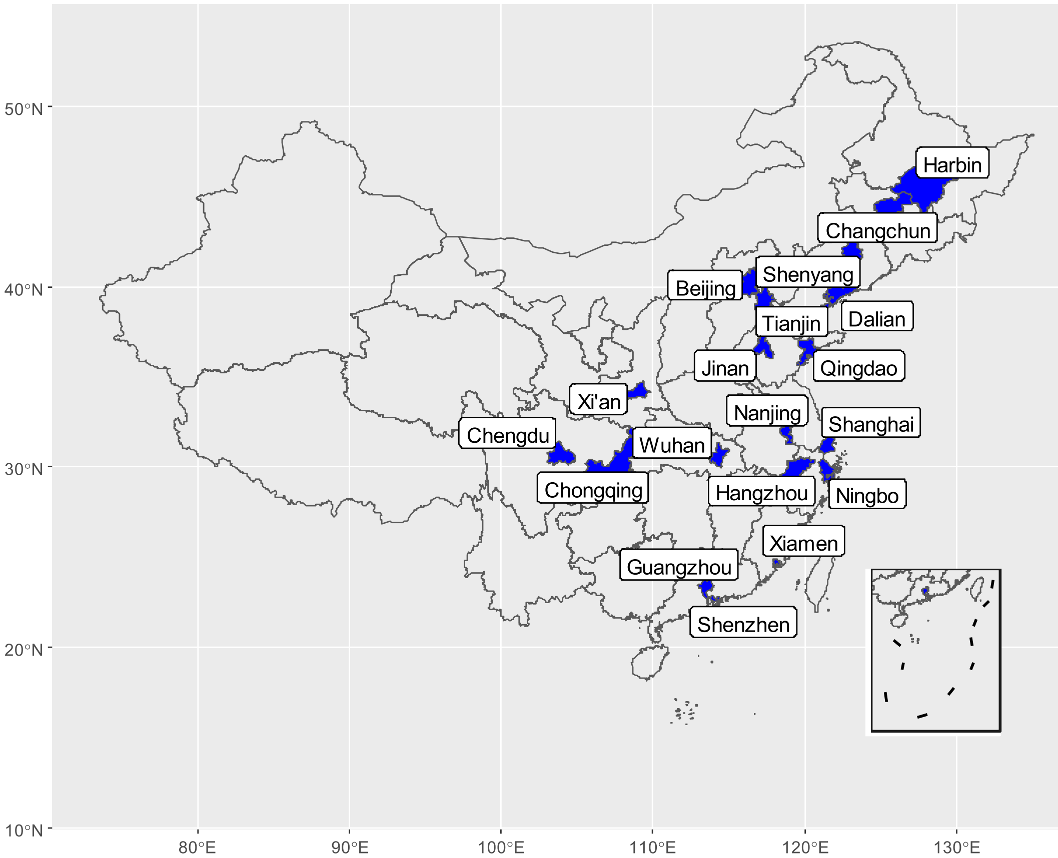


**Fig. 2.** Map of 19 cities.

equation models. Four of the five endogenous variables (i.e., commuting mode, household car ownership, household location, and housing source) are categorical. For these non-normal endogenous variables, diagonally weighted least squares yields more accurate parameter estimates because it makes no distributional assumptions on endogenous variables (Mîndrilă, 2010). The diagonally weighted least squares estimation of the simultaneous equation models was implemented by using the lavaan package in R.

## 4. Data and variables

### 4.1. Data

The data were from the China Household Finance Survey (CHFS), which was conducted by Southwestern University of Finance and Economics in 2017 (Gan et al., 2014). China Household Finance Survey 2017 surveyed 40,011 households with a stratified probability sampling strategy, covering 29 provinces and municipalities. In the China Household Finance Survey 2017 dataset, 9,600 individuals living in four municipalities (i.e., Beijing, Shanghai, Tianjin, and Chongqing) and fifteen sub-provincial cities (i.e., Guangzhou, Wuhan, Harbin, Shenyang, Chengdu, Nanjing, Xi'an, Changchun, Jinan, Hangzhou, Dalian, Qingdao, Shenzhen, Xiamen, and Ningbo) were selected as the sample for this study. The map of the 19 cities is shown in Fig. 2 and the basic data, including population, GDP per capita, vehicle ownership per thousand inhabitants, commuting distance, and commuting time, are presented in Table 1. The municipalities and sub-provincial cities are the two highest administrative levels and the most developed cities in China. The total population of these cities is 182 million, accounting for 13% of China's total population. The urban transportation in these cities has changed dramatically over decades of rapid economic growth and urbanization, with residents traveling longer distances and relying more heavily on motorized vehicles. The mean commuting distance and commuting time of these cities are 8.2 km and 37 min, respectively. In contrast, the mean commuting distance and commuting time of other provincial cities and prefecture-level cities (the two middle administrative levels) are 7.0 km and 31 min, respectively (Baidu Map, 2018).

We divided the 9,600 individuals into the private housing group and the social housing group based on homeownership status and housing source. There were three types of homeownership status (i.e., homeowner, tenant, and lodge for free) in the China Household Finance Survey 2017. If the homeowner was selected, respondents were further required to select among nine types of housing sources (i.e., new commodity housing, second-hand commodity housing, affordable housing, inherited housing, work-unit housing, fund-raising cooperative housing, self-built housing, replacement housing, and limited property housing). If tenant or lodge for free was selected, respondents were further required to select among four types of housing sources (i.e., government-provided, work unit-provided, relative-provided, and non-relative-provided). The private housing group comprised homeowners of new or second-hand commodity housing, tenants of relative- or non-relative-provided housing, and people who lodge for free in relative- or non-relative-provided housing. Individuals in this group obtained their housing on the open market and had a high degree of flexibility in choosing where to live. The rest of the sample formed the social housing group. Individuals in this group had very little flexibility in choosing their residences. Based on this definition, out of the 9,600 individuals, 6,059 individuals were assigned to the private housing group and 3,541 individuals were assigned to the social housing group.

Household locations were divided into urban and suburban neighborhoods, based on information provided directly by the China Household Finance Survey 2017. In large cities in China, urban areas are characterized by a traditional street grid, high accessibility,

**Table 1**
Basic data of 19 cities.

| City | Population (10 thousand) | GDP per capita (Yuan) | Vehicle ownership per thousand inhabitants | Commuting distance (km) | Commuting time (min) |
|---|---|---|---|---|---|
| Beijing | 2,162 | 140,211 | 0.261 | 11.1 | 47 |
| Shanghai | 2,421 | 134,982 | 0.148 | 9.1 | 42 |
| Tianjin | 1,558 | 120,711 | 0.184 | 8.5 | 39 |
| Chongqing | 3,088 | 65,933 | 0.120 | 9.1 | 40 |
| Guangzhou | 1,470 | 155,491 | 0.163 | 8.7 | 38 |
| Wuhan | 1,099 | 135,136 | 0.238 | 8.2 | 38 |
| Harbin | 953 | 66,094 | 0.173 | 7.2 | 35 |
| Shenyang | 831 | 75,766 | 0.253 | 7.4 | 35 |
| Chengdu | 1,619 | 94,782 | 0.279 | 9.1 | 39 |
| Nanjing | 839 | 152,886 | 0.285 | 8.5 | 39 |
| Xi'an | 981 | 85,114 | 0.276 | 8.3 | 35 |
| Changchun | 750 | 95,663 | 0.191 | 7.5 | 35 |
| Jinan | 739 | 106,302 | 0.280 | 8.0 | 34 |
| Hangzhou | 964 | 140,180 | 0.253 | 7.4 | 35 |
| Dalian | 700 | 109,550 | 0.216 | 7.3 | 37 |
| Qingdao | 934 | 128,459 | 0.263 | 8.1 | 39 |
| Shenzhen | 1,278 | 189,568 | 0.252 | 8.1 | 36 |
| Xiamen | 406 | 118,015 | 0.303 | 7.1 | 33 |
| Ningbo | 810 | 132,603 | 0.283 | 6.6 | 32 |
| Mean | 1,242 | 118,287 | 0.233 | 8.2 | 37 |

Source: Population, GDP per capita, and vehicle ownership per thousand inhabitants are from the China City Statistical Yearbook (2019), and commuting distance and commuting time are from China's Major Cities Commuting Monitoring Report (2020).

and high diversity of amenities, while suburban areas are characterized by large blocks and wide streets, low accessibility, and low diversity of amenities. The China Household Finance Survey does not record detailed household locations, but it does record the availability of various public facilities such as schools, banks, and parks, as well as transit facilities, such as metro stations and bus stops within 1 km walking distance of household locations. We used the percentages of households with different public and transit facilities to depict the residential built environment in urban and suburban areas. As shown in Table 2, urban areas tend to have greater access to all public and transit facilities. In the private housing group, 4,514 individuals lived in the urban areas and 1,545 individuals lived in the suburban areas. In the social housing group, 2,458 individuals lived in the urban areas and 1,083 individuals lived in the suburban areas.

### 4.2. Variables

We used six sociodemographic control variables (i.e., gender, age, marital status, education level, income, and household size), which are similar to the sociodemographic control variables used in previous studies (Yang et al., 2017; Deng and Yan, 2019; van Herick and Mokhtarian, 2020). Gender is a nominal variable with two options (female and male); marital status is a nominal variable with two options (single and married); education level is a quantitative variable ranging from 1 to 9 representing never attended school, primary school, secondary school, high school, polytechnic, junior college, university undergraduate, master's, and doctorate, respectively; age, income, and household size are quantitative variables. These sociodemographic control variables were used not only as covariates in propensity score weighting to eliminate the sociodemographic differences among the four groups, but they were also included in the simultaneous equation models as additional covariate adjustment to obtain doubly robust estimates.

Household location, housing source, household car ownership, and commuting distance are mediating endogenous variables. Household location is represented by a dummy variable that has a value of 1 if the individual lives in a suburban neighborhood and 0 if the individual lives in an urban neighborhood. Housing source is represented by a dummy variable that has a value of 1 if the individual lives in private housing and 0 if the individual lives social housing. Since the sociodemographic differences among the four groups were eliminated by propensity score weighting, the effect of household location can be attributed to the built environment and the effect of housing source can be attributed to residential self-selection. Note that the effects of the built environment and residential self-selection we estimated is the modular effects as termed by van Herick and Mokhtarian (2020, 2021), which is one of the three methods for estimating the built environment proportion as well as the effect of residential self-selection. Household car ownership and commuting distance were directly included in the simultaneous equation models since they are quantitative variables.

The final endogenous variable is the binary commuting mode. The original commuting mode has eight options (i.e., bus, rail transit, company bus, car, taxi, e-bike, bike, and walking). We focused on five major commuting modes: public transit (including bus, rail transit, company bus, and taxi), car, regular bike, e-bike, and walking. Each mode was represented by a binary commuting mode variable. In the dataset, twenty-one percent of the respondents drive to work, 34% take public transit to commute to work, 9% commute by regular bike, 15% commute by e-bike, and 21% commute by foot.

## 5. Results

### 5.1. Results of propensity score weighting

Table 3 shows the mean values of the control variables for each group before and after weighting. Before weighting, private housing residents were more likely to be female, married, younger, better educated, and having higher incomes and larger households compared to social housing residents. Urban residents were more likely to be unmarried, better educated, and having higher incomes and small households compared to suburban residents. After weighting, the four groups were almost matched. Their sample sizes also become larger and closer because each observation was assigned a weight greater than 1. We used pairwise absolute standardized mean differences (ASMD) and a pairwise *t*-test to assess the balance of each covariate. Small values of absolute standardized mean difference ($<$0.1) and large p-values of t-tests ($>$0.1) are expected if the balance has been achieved. The maximum values of pairwise absolute standardized mean difference for the covariates get smaller, with values decreasing from 0.106 to 0.756 before weighting to 0.020 to 0.108 after weighting. The minimum p-values of pairwise t-tests show that, before weighting, five covariates were significant at the 0.001 level and one covariate was significant at the 0.01 level, while after weighting, four covariates were significant at the 0.1 level and two covariates were insignificant. This result verifies the use of additional covariate adjustment in the simultaneous equation models. The propensity score weighting has made groups more similar, perhaps enough so that additional modeling with covariates can adjust for any remaining differences.

Notably different patterns of household car ownership, commuting distance, and commuting mode shares can be observed in Table 3. Urban residents generally have more public transit, regular bike, and walking mode shares and fewer car and e-bike mode shares than suburban residents for both private housing and social housing groups. Social housing residents generally have more

**Table 2**
Accessibility of facilities for urban and suburban residents.

| | Metro station | Bus stop | Kindergarten | Primary school | Middle school | Park | Bank | Hospital |
|---|---|---|---|---|---|---|---|---|
| Urban | 31% | 95% | 77% | 78% | 62% | 60% | 86% | 72% |
| Suburban | 10% | 81% | 63% | 66% | 44% | 41% | 65% | 57% |

**Table 3**
Descriptive statistics for variables before and after weighting.

| Variable | | Private housing | | Social housing | | Maximum value of pairwise ASMD | Minimum p-value of pairwise *t*-test |
|---|---|---|---|---|---|---|---|
| | | Suburban | Urban | Suburban | Urban | | |
| Gender | before | 0.465 | 0.456 | 0.413 | 0.425 | 0.106 | 0.008** |
| (Ref. male) | after | 0.447 | 0.449 | 0.436 | 0.447 | 0.027 | 0.524 |
| Age | before | 40.01 | 39.48 | 42.56 | 42.93 | 0.303 | 0.000*** |
| | after | 40.62 | 40.69 | 41.07 | 41.07 | 0.040 | 0.268 |
| Marital status | before | 0.867 | 0.811 | 0.815 | 0.747 | 0.302 | 0.000*** |
| (Ref. single) | after | 0.822 | 0.806 | 0.767 | 0.804 | 0.108 | 0.013* |
| Education level | before | 4.570 | 5.538 | 4.152 | 5.081 | 0.756 | 0.000*** |
| | after | 5.017 | 5.134 | 4.961 | 5.108 | 0.094 | 0.032* |
| Income (in 1,000 Yuan) | before | 52.73 | 71.61 | 39.06 | 50.10 | 0.432 | 0.000*** |
| | after | 56.52 | 59.71 | 53.10 | 57.57 | 0.089 | 0.026* |
| Household size | before | 3.483 | 3.125 | 3.774 | 3.245 | 0.524 | 0.000*** |
| | after | 3.278 | 3.259 | 3.362 | 3.291 | 0.083 | 0.038* |
| Household car ownership | before | 0.607 | 0.590 | 0.554 | 0.450 | | |
| | after | 0.617 | 0.533 | 0.628 | 0.505 | | |
| Commuting distance (km) | before | 8.124 | 9.858 | 7.534 | 8.611 | | |
| | after | 9.031 | 8.965 | 8.680 | 8.731 | | |
| Car mode | before | 0.249 | 0.237 | 0.169 | 0.146 | | |
| | after | 0.265 | 0.204 | 0.249 | 0.171 | | |
| Public transit mode | before | 0.236 | 0.392 | 0.187 | 0.375 | | |
| | after | 0.273 | 0.367 | 0.226 | 0.363 | | |
| Regular bike mode | before | 0.076 | 0.079 | 0.106 | 0.110 | | |
| | after | 0.069 | 0.089 | 0.099 | 0.103 | | |
| E-bike mode | before | 0.212 | 0.108 | 0.313 | 0.136 | | |
| | after | 0.185 | 0.125 | 0.243 | 0.139 | | |
| Walking mode | before | 0.227 | 0.184 | 0.224 | 0.232 | | |
| | after | 0.208 | 0.216 | 0.183 | 0.224 | | |
| N | before | 4514 | 1545 | 2458 | 1083 | | |
| | after | 9283 | 9033 | 8869 | 8029 | | |

***, **, * denote significance at the 0.001, 0.01, and 0.1 levels, respectively.

regular bike and e-bike mode shares but fewer car and public transit mode shares than private housing residents for both urban and suburban groups.

### 5.2. Results of simultaneous equation models

Because all variables are directly observed (without any latent variables) and every variable has a free relation with every other variable, the five simultaneous equation models are just-identified (i.e., have zero degrees of freedom), which renders typical goodness-of-fit measures (e.g., CFI, TLI, RMSEA, and SRMR) meaningless since they all show a perfect fit. We instead used McElroy $R^2$ to indicate the overall model fit. McElroy $R^2$ ranges from 0.09 to 0.15 for these models. McElroy $R^2$ is not very high because the predicted values of endogenous variables (i.e., housing source, household location, car ownership, and commuting mode) could fall outside the range of 0 and 1 despite the observed values being 0 or 1.

Table 4 shows the estimation results of the direct effects of these models. A direct effect is the effect of a variable on another variable without any mediating variable. The direct effects of housing source and household location on the mediating variables and the direct effects of the mediating variables on commuting mode choice are most as expected. The only exception is that the direct effect of housing source on household location is 0.020, indicating that private housing residents who can self-select their place of residence tend to live in suburban areas. We discuss the reasons in the next section. In terms of the effects on the mediating variables, suburban residents own more cars and have longer commuting distances than urban residents. In terms of the effects of the mediating variables, longer commuting distances increase the shares of car and public transit commuting modes and decrease those of regular bike, e-bike, and walking commuting modes. More household car ownership is associated with more commuting by the car mode and less commuting by all other modes. Note that the regression coefficients for the four endogenous variables (i.e., household car ownership, commuting distance, household location, and housing source) are the same in all simultaneous equation models because the regression equations for these variables are the same in all simultaneous equation models.

Table 5 shows the direct, indirect, and total effects of household location and housing source on commuting mode choice. An indirect effect is the effect exerted by a variable on another variable through any mediating variable, which can be calculated by multiplying the effect on the mediator and the effect of the mediator. A total effect is the sum of the direct and indirect effects. The significance of both indirect and total effects is tested by bootstrapping. In general, the direct effects of household location and housing source are larger in magnitude than their indirect effects. Most of the direct and indirect effects are in the same directions, except for the effects of household locations on e-bike mode and housing source on walking mode. Thus, most of the total effects of household location and housing source are larger in magnitude than their direct effects.

Fig. 3 shows the path diagrams. Household location influences commuting mode choice through four paths: direct, indirect through

**Table 4**
Estimation results of simultaneous equation models.

| Variable | Housing source | Household location | Commuting distance (in 10 km) | Household car ownership | Model 1: Car mode | Model 2: Transit mode | Model 3: Regular bike mode | Model 4: E-bike mode | Model 5: Walking mode |
|---|---|---|---|---|---|---|---|---|---|
| Gender (Ref. male) | 0.004 | −0.009 | −0.098*** | 0.028* | −0.120*** | 0.080*** | 0.009 | −0.029** | 0.052*** |
| Age (in 10 years) | −0.056** | −0.007 | −0.009 | 0.010 | −0.031* | −0.035 | 0.046* | −0.027 | 0.065** |
| Marital status (Ref. single) | 0.069*** | −0.019 | −0.001 | 0.158*** | 0.041* | −0.079*** | 0.022 | 0.067*** | −0.029 |
| Education level | −0.006 | −0.037** | 0.160*** | 0.268*** | 0.065*** | 0.144*** | −0.015 | −0.205*** | −0.031 |
| Income (in 1,000 Yuan) | 0.013 | −0.015 | 0.082*** | 0.165*** | 0.111*** | −0.044* | −0.006 | −0.020* | −0.040** |
| Household size | −0.046*** | 0.021 | 0.051*** | 0.169*** | −0.035* | 0.009 | −0.035* | 0.049*** | 0.005 |
| Housing source | | 0.020* | 0.011 | −0.001 | 0.020* | 0.024* | −0.035*** | −0.044*** | 0.017 |
| Household location | | | 0.007* | 0.093*** | 0.052*** | −0.102*** | −0.014 | 0.107*** | −0.026** |
| Commuting distance (in 10 km) | | | | 0.023* | 0.086*** | 0.347*** | −0.127*** | −0.084*** | −0.316*** |
| Household car ownership | | | | | 0.470*** | −0.231*** | −0.081*** | −0.107*** | −0.062*** |
| McElroy $R^2$ | | | | | 0.15 | 0.13 | 0.10 | 0.09 | 0.11 |

***, **, * denote significance at the 0.001, 0.01, and 0.1 levels, respectively.

**Table 5**
The direct, indirect, and total effects on commuting modes.

| Variable | Car mode | | | Transit mode | | | Regular bike mode | | | E-bike mode | | | Walking mode | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total |
| Housing source | 0.020* | 0.001 | 0.021* | 0.024* | 0.004 | 0.028** | −0.035*** | −0.001 | −0.037*** | −0.044*** | −0.001 | −0.045*** | 0.017 | −0.004 | 0.014 |
| Household location | 0.052*** | 0.044*** | 0.096*** | −0.102*** | −0.019*** | −0.121*** | −0.014 | −0.008*** | −0.023* | 0.107*** | −0.011*** | 0.097*** | −0.026** | −0.008* | −0.034*** |

***, **, * denote significance at the 0.001, 0.01, and 0.1 levels, respectively.

commuting distance, indirect through household car ownership, and indirect through both commuting and household car ownership. Housing source influences commuting mode choice not only through the same four paths, but also by influencing household location first and then through the previous four paths. The total effects of household location on car, public transit, regular bike, e-bike, and walking modes are 0.096, −0.121, −0.023, 0.097, and −0.034, respectively, indicating an increase in car and e-bike mode shares and a decrease in public transit, regular bike and walking mode shares in suburban areas compared to urban areas. The total effects of housing source on car, public transit, regular bike, and e-bike modes are 0.021, 0.028, −0.037, and −0.045, respectively, indicating that the private housing residents who can self-select their place of residence tend to commute by car or public transit, but not by regular bike or e-bike.

## 6. Discussion

### *6.1. Empirical findings of residential self-selection*

In this section, we discuss the relationship between household location and housing source, and their total effects on commuting mode choice. The result that private housing residents who can self-select their place of residence tend to live in suburban areas is different from common sense in the large cities of China. We offer two possible reasons. First, people have different travel preferences, and some who prefer car commuting may choose to live in suburban areas because of less congested roads. Second, household location choices are affected not only by travel preferences but also by preferences for dwelling and neighborhood qualities that are not related to travel (Wolday et al., 2019). For the same cost, people living in suburban areas can have larger dwelling sizes and better neighborhoods than those living in urban areas. Note that the existence of non-travel-related residential preferences may bias the estimates of residential self-selection. However, we believe that the possible bias caused by non-travel-related residential preferences is limited

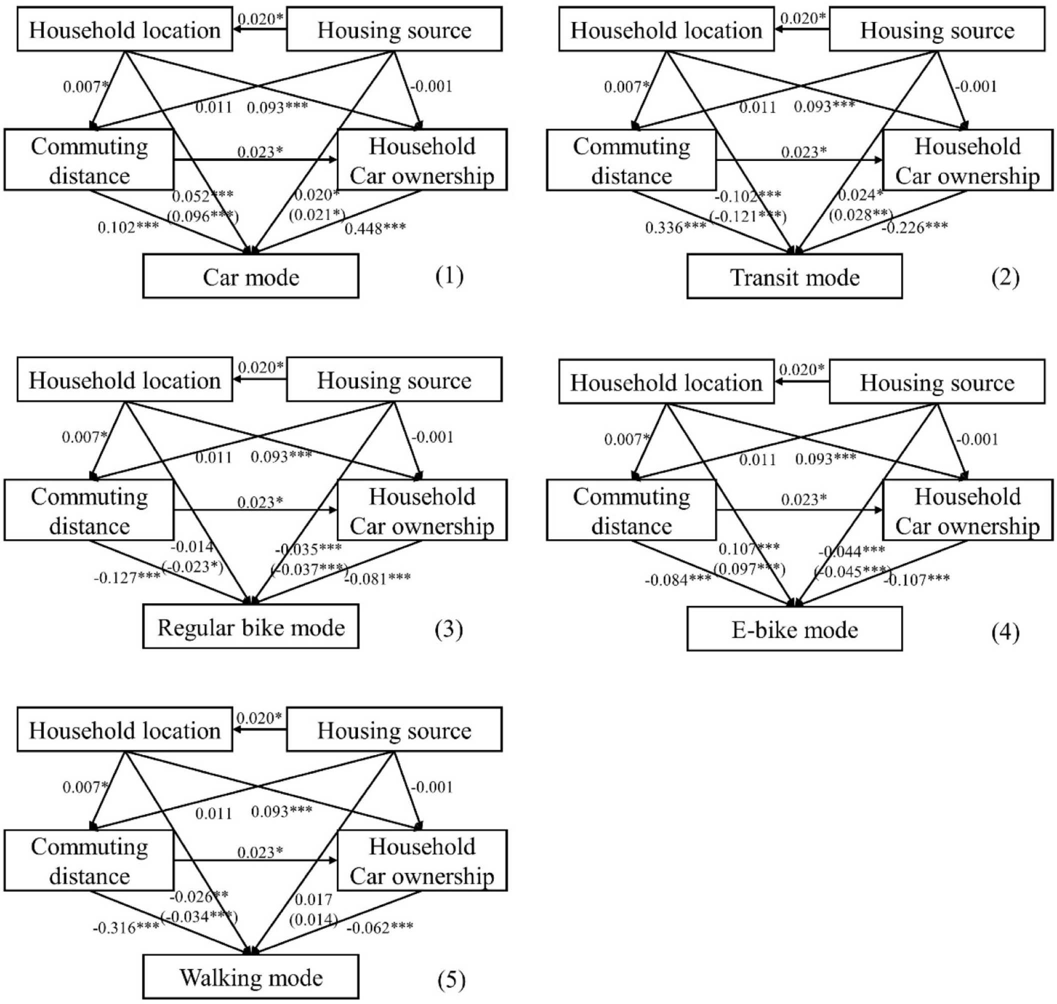


**Fig. 3.** Path diagrams of simultaneous equation models. Note: Numbers in parentheses are total effects. ***, **, * denote significance at the 0.001, 0.01, and 0.1 levels, respectively.

in this study because the magnitude of the effect of housing source via household location on commuting mode choice is much smaller than either the total effects of household location or housing source on commuting mode choice.

Since the effect of household location can be attributed to the built environment, and the effect of housing source can be attributed to residential self-selection, we calculated the built environment proportions for different commuting modes and discussed the relative magnitude of residential self-selection. Built environment proportion represents the proportion of the effect of the built environment to the total effects of the built environment and residential self-selection. This proportion was obtained by dividing the value of the coefficient of the built environment by the sum of the values of the coefficients of the built environment and residential self-selection. The built environment proportions of car, public transit, regular bike, and e-bike modes are 0.821, 1.301, 0.383, and 1.865, respectively. Proportions<100% imply the existence of residential self-selection (most people live in preferred residence), with lower proportions indicating a more important impact of residential self-selection; proportions> 100% imply the existence of residential dissonance (most people live in undesired residence), and the true impact of the built environment is even greater than the observed impact. We did not calculate the built environment proportion of the walking mode because the effect of residential self-selection is insignificant. Note that the residential self-selection effect we calculated does not included the effect induced by socio-demographics.

For the car mode, the coefficient signs of the built environment and residential self-selection are the same. Private housing residents who can self-select their residential locations prefer to live in suburban neighborhoods and commute by car. The built environment proportion (82.1%) shows the built environment proportion has a relatively larger influence than residential self-selection for the car mode, which consistent with most previous studies (Mokhtarian and Van Herick, 2016).

Residential self-selection has the opposite direction of impact as compared to the built environment for the public transit mode and the built environment proportion is large (130.1%). Private housing residents who can self-select their place of residence prefer to commute by transit but live in suburban neighborhoods, where public transit is not well served. In contrast, a study in the United States found that residential self-selection and the built environment impact public transit commuting in the same direction (Lee et al., 2014). In China, good accessibility to public transit is a common consideration in residential choices, as most people do not own private cars. Whether people choose neighborhoods with good public transit is driven primarily by economic considerations rather than attitudes towards public transit (Wang and Lin, 2014). In contrast, whether people in the United States choose neighborhoods with good public transit is influenced by both economic constraints and attitudes towards public transit (Mokhtarian and Cao, 2008). Thus the travel preferences of public transit are not important for Chinese people in choosing their place of residence, which may explain the opposite sign of residential self-selection in China.

For the regular bike mode, the coefficient signs of the built environment and residential self-selection are the same, while for the e-bike mode, the coefficient signs are opposite. Few studies have focused on the impact of residential self-selection on bike commuting, especially e-bike commuting. We suggest the following reason for the dissonance of the e-bike mode. People who can self-select their place of residence may have both a travel preference for not commuting by e-bike and a residential preference for suburban areas. However, these two preferences are in conflict in the choice of residence, and the travel preference may give way to the residential preference because the travel preference is not the dominant criterion in household choice (Wee and Levinson, 2014). In contrast to other transportation modes that are more likely to be constrained by the built environment, especially transportation facilities such as bus stops, the e-bike mode is less constrained by transportation facilities. Commuting by e-bike seems to be a forced choice for suburban residents. In addition, the built environment has opposite effects on the regular bike mode and e-bike mode. This result confirms the findings of Deng and Yan' study in China that a good residential built environment may encourage residents to ride regular bikes, while a bad one may force residents to ride e-bikes to reach farther distances (Deng and Yan, 2019). Our study confirms the need to separate regular bike and e-bike modes when studying the impact of the built environment and residential self-selection.

### 6.2. *Justification of the methodology*

The methodological framework we proposed has the advantages of distinguishing the housing source and making the estimates of built environment proportions more reliable by the application of a natural experiment and doubly robust estimation. In this section, we discuss the differences in results with and without using the natural experiment and doubly robust estimation.

When assessing the effects of the built environment and residential self-selection, previous studies have usually only divided residents into suburban and urban groups, performed propensity score matching to correct for sociodemographic and attitudinal heterogeneity, and then conducted difference-in-means type analysis (Yang et al., 2017; Cheng et al., 2019; Deng and Yan, 2019). These studies have rarely distinguished between private household residents and social housing residents in that the latter are treated as same as the former. For social housing residents, the effect of the built environment is partially misattributed to the effect of residential self-selection that should not exist, which leads to an underestimation of built environment proportion. Therefore, it is theoretically important to distinguish between social housing residents and private housing residents in societies with large-scale social housing programs.

Previous studies usually performed single robust estimation methods (propensity score matching or statistical control) to estimate the effects of the built environment and residential self-selection, while in this study, we used the doubly robust estimation method (simultaneous equation models with covariates after propensity score weighting). Theoretically, the doubly robust estimation method allows us to obtain more robust estimates of the effects of the built environment and residential self-selection because it minimizes the mean square error by addition propensity score adjustment or additional covariate adjustment, which may make the estimates more significant and reliable.

We further compared the differences between the doubly robust estimates and single robust estimates empirically. The doubly robust estimates were obtained in the previous calculations. In calculating the single robust estimates, we used the same data and

**Table 6**
Results of single and doubly robust estimations.

| Variable | Control method | Car | Public transit | Regular bike | E-bike | Walking |
|---|---|---|---|---|---|---|
| Built environment | doubly robust | 0.096 (0.008) *** | −0.121 (0.008) *** | −0.023 (0.005) * | 0.097 (0.008) *** | −0.034 (0.008) ** |
| | single robust I | 0.066 (0.011) *** | −0.123 (0.012) *** | −0.014 (0.007) | 0.099 (0.011) *** | 0.000 (0.011) |
| | single robust II | 0.083 (0.008) *** | −0.123 (0.008) *** | −0.021 (0.005) * | 0.107 (0.008) *** | −0.029 (0.008) ** |
| Residential self-selection | doubly robust | 0.021 (0.008) * | 0.028 (0.009) * | −0.037 (0.006) *** | −0.045 (0.008) *** | 0.014 (0.008) |
| | single robust I | 0.056 (0.010) *** | 0.017 (0.012) | −0.037 (0.007) ** | −0.041 (0.009) *** | −0.013 (0.010) |
| | single robust II | 0.030 (0.008) *** | 0.026 (0.009) * | −0.037 (0.006) *** | −0.047 (0.008) *** | 0.009 (0.008) |
| Built environment proportion | doubly robust | 0.821 | 1.301 | 0.383 | 1.865 | — |
| | single robust I | 0.541 | — | — | 1.707 | — |
| | single robust II | 0.735 | 1.268 | 0.362 | 1.783 | — |

Note: Single robust I refers to only covariate adjustment and single robust II refers to only propensity score adjustment. The values in parentheses are standard errors. ***, **, * denote significance at the 0.001, 0.01, and 0.1 levels, respectively. — denotes that the built environment proportion is not calculated because either the built environment or residential self-selection is insignificant.

models, with the only difference being that propensity score weighting was no longer implemented (single robust estimates type I) or that the control variables were no longer included in the simultaneous equation models (single robust estimates type II). This calculation process can ensure that the difference in results is only due to whether the doubly robust estimation was performed. Table 6 shows the results of single and doubly robust estimations. The single robust estimates type I, which has only covariate adjustment, causes increases in standard errors and two effects even become insignificant. The changes in the effects of the built environment and residential self-selection further lead to changes in the built environment proportions. The single robust estimates type II, which only has propensity score adjustment, does not change standard errors but slightly changes the effects of the built environment and residential self-selection. The resulting changes in the built environment proportions are small (<0.1). Therefore, for a natural experimental design, covariate adjustment alone is not sufficient to exclude sociodemographic heterogeneity, especially when the significance levels of the effects of the built environment and residential self-selection are low. Even with propensity score adjustment, additional covariate adjustment is still recommended because half of the sociodemographic control variables are still significant in the simultaneous equation models though propensity score weighting was performed (as shown in Table 4).

### 6.3. *Justification of the results*

We performed a sensitivity analysis to assess the robustness of the results. We reran the models with a sample of 3,592 people from four municipalities (i.e., Beijing, Shanghai, Tianjin, and Chongqing). The results are shown in Table 7 and Table 8. Except for the effect of commuting distance on household car ownership that becomes insignificant, the coefficients of the new models change only slightly in magnitude compared to the coefficients of the original models. Commuting distance has an insignificant effect on household car ownership for two reasons: 1) all these municipalities have well-served public transit systems, especially urban rail transit and 2) all these municipalities restrict some vehicles from driving on weekdays based on license plate numbers, and three of them have adopted

**Table 7**
Estimation results of simultaneous equation models for a sample of four municipalities.

| Variable | Housing source | Household location | Commuting distance (in 10 km) | Household car ownership | Model 1: Car mode | Model 2: Transit mode | Model 3: Regular bike mode | Model 4: E-bike mode | Model 5: Walking mode |
|---|---|---|---|---|---|---|---|---|---|
| Gender (Ref. male) | 0.002 | −0.009 | −0.097*** | 0.040* | −0.113*** | 0.086*** | −0.014 | −0.018 | 0.043* |
| Age (in 10 years) | −0.093** | 0.006 | 0.030 | 0.030 | −0.005 | −0.053 | 0.053 | 0.036 | −0.014 |
| Marital status (Ref. single) | 0.098*** | −0.018 | −0.003 | 0.152*** | 0.007 | −0.086** | 0.049* | 0.063* | 0.003 |
| Education level | −0.019 | −0.042* | 0.180*** | 0.219*** | 0.039 | 0.190*** | −0.032 | −0.194*** | −0.078** |
| Income (in 1,000 Yuan) | 0.023 | −0.037* | 0.138** | 0.182** | 0.132* | −0.072* | −0.015 | −0.022 | −0.023 |
| Household size | −0.058** | 0.036* | 0.095*** | 0.164*** | −0.007 | 0.005 | −0.046* | 0.059* | −0.011 |
| Housing source | | 0.032* | 0.014 | −0.011 | 0.027* | 0.028* | −0.024* | −0.047** | −0.003 |
| Household location | | | 0.020* | 0.174*** | 0.056** | −0.107*** | −0.008 | 0.127*** | −0.031* |
| Commuting distance (in 10 km) | | | | 0.015 | 0.104*** | 0.342*** | −0.165*** | −0.083*** | −0.318*** |
| Household car ownership | | | | | 0.469*** | −0.253*** | −0.091*** | −0.108*** | −0.029* |
| McElroy $R^2$ | | | | | 0.16 | 0.14 | 0.10 | 0.11 | 0.11 |

***, **, * denote significance at the 0.001, 0.01, and 0.1 levels, respectively.

**Table 8**
The direct, indirect, and total effects on commuting modes for a sample of four municipalities.

| Variable | Car mode | | | Transit mode | | | Regular bike mode | | | E-bike mode | | | Walking mode | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total |
| Housing source | 0.027* | −0.001 | 0.026* | 0.028* | 0.007 | 0.035* | −0.024* | −0.002 | −0.026* | −0.047** | −0.001 | −0.048** | 0.003 | −0.005 | 0.008 |
| Household location | 0.056** | 0.084*** | 0.139*** | −0.107*** | −0.037*** | −0.144*** | −0.008 | −0.018*** | −0.026* | 0.127*** | −0.021*** | 0.106*** | −0.031* | −0.011* | −0.043** |

***, **, * denote significance at the 0.001, 0.01, and 0.1 levels, respectively.

car ownership policies of license plate auctions and license plate lotteries.

## 7. Conclusions

Estimating the influence of the built environment and residential self-selection on travel behavior has been a topic of policy and academic interest for decades. In this study, we proposed a natural experiment framework that incorporates propensity score weighting and simultaneous equation models and used multi-city data to reveal the influence of the built environment and residential self-selection on commuting mode choices. We found that the built environment and residential self-selection have direct and indirect effects through commuting distance and household car ownership on commuting mode choice. The total effects of the built environment and residential self-selection are in opposite directions for public transit and e-bike modes. Key policy implications include 1) Improving the built environment can lead to positive changes in commuting mode choice (i.e., increasing the public transit, regular bike and walking mode shares and decreasing the car and e-bike mode shares), and thus reduce the negative impacts of commuting such as traffic congestion and air pollution. 2) Land use policies would achieve additional success in increasing public transit commuting if they can enable people to self-select their household locations. 3) Transportation policies should properly direct e-bike mode, as e-bike is an important commuting mode for suburban residents, yet residents do not like e-bike.

This study also has some limitations. First, the China Household Finance Survey classified the residential built environment into urban and suburban areas, and this simple classification may not be sufficient to describe the differences in the built environment. Second, we determined whether people have the flexibility to choose the place of residence based on their housing source, which is not always true. For example, it is possible that some people living in private housing have limited housing choices due to the high real estate prices in the large cities of China. Third, we used the residential self-selection framework, ignoring the possible effects of the built environment on travel attitudes over time. Future studies could use qualitative methods to distinguish whether people have the flexibility to choose their residential locations and to determine whether the residential self-selection or the residential determination framework is more plausible.

## CRediT authorship contribution statement

**Yiling Deng:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Writing – original draft. **Pengjun Zhao:** Conceptualization, Funding acquisition, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Acknowledgments*

This work was supported by the National Natural Science Foundation of China [grant numbers 51978617 and 41925003].

## References

Baidu Map, 2018. China urban transportation report 2018.
Ministry of Housing and Urban-Rural Development of China, 2019. China city statistical yearbook.
Ministry of Housing and Urban-Rural Development of China, 2020. China’s major cities commuting monitoring report 2020.
Ababio-Donkor, A., Saleh, W., Fonzone, A., 2020. Understanding transport mode choice for commuting: The role of affect. Transport. Plan. Techn. 43 (4), 385–403. https://doi.org/10.1080/03081060.2020.1747203.
Bautista-Hernández, D.A., 2021. Mode choice in commuting and the built environment in México City. Is there a chance for non-motorized travel? J. Transp. Geogr. 92, 103024. https://doi.org/10.1016/j.jtrangeo.2021.103024.
Bhat, C.R., Eluru, N., 2009. A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transp. Res. Part B Methodol. 43 (7), 749–765. https://doi.org/10.1016/j.trb.2009.02.001.
Bohte, W., Maat, K., van Wee, B., 2009. Measuring attitudes in research on residential Self-Selection and travel behaviour: A review of theories and empirical research. Transp. Rev. 29 (3), 325–357. https://doi.org/10.1080/01441640902808441.

Cao, J., Ettema, D., 2014. Satisfaction with travel and residential self-selection: How do preferences moderate the impact of the Hiawatha Light Rail Transit line? J. Transp. Land Use 7 (3), 93–108. https://doi.org/10.5198/jtlu.v7i3.485.

Cao, X., 2015. Examining the impacts of neighborhood design and residential self-selection on active travel: A methodological assessment. Urban Geogr. 36 (2), 236–255. https://doi.org/10.1080/02723638.2014.956420.

Cao, X., Yang, W., 2017. Examining the effects of the built environment and residential self-selection on commuting trips and the related $CO_2$ emissions: An empirical study in Guangzhou, China. Transp. Res. Part D Transp. Environ. 52, 480–494. https://doi.org/10.1016/j.trd.2017.02.003.

Cao, X.J., 2009. Disentangling the influence of neighborhood type and self-selection on driving behavior: An application of sample selection model. Transportation 36 (2), 207–222. https://doi.org/10.1007/s11116-009-9189-9.

Cao, X.J., Mokhtarian, P.L., Handy, S.L., 2009. Examining the Impacts of Residential Self-Selection on Travel Behaviour: A Focus on Empirical Findings. Transp. Rev. 29 (3), 359–395. https://doi.org/10.1080/01441640802539195.

Cheng, L., De Vos, J., Shi, K., Yang, M., Chen, X., Witlox, F., 2019. Do residential location effects on travel behavior differ between the elderly and younger adults? Transp. Res. Part D Transp. Environ. 73, 367–380. https://doi.org/10.1016/j.trd.2019.07.015.

Clark, B., Chatterjee, K., Melia, S., 2016. Changes to commute mode: The role of life events, spatial context and environmental attitude. Transp. Res. Part A Policy Pract. 89, 89–105. https://doi.org/10.1016/j.tra.2016.05.005.

Deng, Y., Yan, Y., 2019. Propensity score weighting with generalized boosted models to explore the effects of the built environment and residential Self-Selection on travel behavior. Transp. Res. Rec. J. Transp. Res. Board. 2673 (4), 373–383. https://doi.org/10.1177/0361198119837153.

Ding, C., Wang, D., Liu, C., Zhang, Y., Yang, J., 2017. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. Transp. Res. Part A Policy Pract. 100, 65–80. https://doi.org/10.1016/j.tra.2017.04.008.

Ettema, D., Nieuwenhuis, R., 2017. Residential self-selection and travel behaviour: What are the effects of attitudes, reasons for location choice and the built environment? J. Transp. Geogr. 59, 146–155. https://doi.org/10.1016/j.jtrangeo.2017.01.009.

Ewing, R., Hamidi, S., Grace, J.B., 2016. Compact development and VMT—Environmental determinism, self-selection, or some of both? Environ. Plann. B Plann. Des. 43 (4), 737–755. https://doi.org/10.1177/0265813515594811.

Gan, L.i., Yin, Z., Jia, N., Xu, S., Ma, S., Zheng, L.u. (Eds.), 2014. Data you need to know about China. Springer Berlin Heidelberg, Berlin, Heidelberg.

Guan, X., Wang, D., 2019. Residential self-selection in the built environment-travel behavior connection: Whose self-selection? Transp. Res. Part D Transp. Environ. 67, 16–32. https://doi.org/10.1016/j.trd.2018.10.015.

Guan, X., Wang, D., 2020. The multiplicity of self-selection: What do travel attitudes influence first, residential location or work place? J. Transp. Geogr. 87, 102809. https://doi.org/10.1016/j.jtrangeo.2020.102809.

Guan, X., Wang, D., Jason Cao, X., 2020. The role of residential self-selection in land use-travel research: A review of recent findings. Transp. Rev. 40 (3), 267–287. https://doi.org/10.1080/01441647.2019.1692965.

Haybatollahi, M., Czepkiewicz, M., Laatikainen, T., Kyttä, M., 2015. Neighbourhood preferences, active travel behaviour, and built environment: An exploratory study. Transp. Res. Part F Traffic Psychol. Behav. 29, 57–69. https://doi.org/10.1016/j.trf.2015.01.001.

Humphreys, J., Ahern, A., 2019. Is travel based residential self-selection a significant influence in modal choice and household location decisions? Transp. Policy 75, 150–160. https://doi.org/10.1016/j.tranpol.2017.04.002.

Lee, B.K., Lessler, J., Stuart, E.A., 2009. Improving propensity score weighting using machine learning. Stat. Med. 29 (3), 337–346. https://doi.org/10.1002/sim.3782.

Lee, J.S., Christopher Zegras, P., Ben-Joseph, E., Park, S., 2014. Does urban living influence baby boomers' travel behavior? J. Transp. Geogr. 35, 21–29. https://doi.org/10.1016/j.jtrangeo.2014.01.004.

Lee, Y., Guhathakurta, S., 2018. An analysis of the effects of suburban densification on vehicle use for shopping: Do existing residents respond to land-use changes in the same way as recent movers? J. Transp. Geogr. 68, 193–204. https://doi.org/10.1016/j.jtrangeo.2018.03.011.

Lin, T., Wang, D., Guan, X., 2017. The built environment, travel attitude, and travel behavior: Residential self-selection or residential determination? J. Transp. Geogr. 65, 111–122. https://doi.org/10.1016/j.jtrangeo.2017.10.004.

Liu, C., Sun, Y., Chen, Y., Susilo, Y.O., 2018. The effect of residential housing policy on car ownership and trip chaining behaviour in Hangzhou, China. Transp. Res. Part D Transp. Environ. 62, 125–138. https://doi.org/10.1016/j.trd.2018.02.008.

Mîndrilă, D., 2010. Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. Int. J. Digit. Soc. 1 (1), 60–66. https://doi.org/10.20533/ijds.2040.2570.2010.0010.

Mokhtarian, P.L., Cao, X., 2008. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. Transp. Res. Part B Methodol. 42 (3), 204–228. https://doi.org/10.1016/j.trb.2007.07.006.

Mokhtarian, P.L., Van Herick, D., 2016. Quantifying residential self-selection effects: A review of methods and findings from applications of propensity score and sample selection approaches. J. Transp. Land Use. 9 (1), 9–28. https://doi.org/10.5198/jtlu.2016.788.

Sun, B., Ermagun, A., Dan, B., 2017. Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai. Transp. Res. Part D Transp. Environ. 52, 441–453. https://doi.org/10.1016/j.trd.2016.06.001.

Van Acker, V., Witlox, F., 2011. Commuting trips within tours: How is commuting related to land use? Transportation 38 (3), 465–486. https://doi.org/10.1007/s11116-010-9309-6.

van Herick, D., Mokhtarian, P.L., 2020. How much does the method matter? An empirical comparison of ways to quantify the influence of residential self-selection. Travel Behav. Soc. 18, 68–82. https://doi.org/10.1016/j.tbs.2019.09.002.

van Herick, D., Mokhtarian, P.L., 2021. Quantifying the share of total apparent effect that is genuinely due to a treatment: A comparison of methods. Res. Transp. Econ. 90, 101020. https://doi.org/10.1016/j.retrec.2020.101020.

van Wee, B., 2009. Self-Selection: A key to a better understanding of location choices, travel behaviour and transport externalities? Transp. Rev. 29 (3), 279–292. https://doi.org/10.1080/01441640902752961.

Wang, D., Lin, T., 2014. Residential self-selection, built environment, and travel behavior in the Chinese context. J. Transp. Land Use. 7 (3), 5–14. https://doi.org/10.5198/jtlu.v7i3.486.

Wee, B.V., Boarnet, M., 2014. Reaction to the paper Tempest in a Teapot: The exaggerated problem of transport-related residential self-selection as a source of error in empirical studies. J. Transp. Land Use. 7 (3), 81. https://doi.org/10.5198/jtlu.v7i310.5198/jtlu.v7i3.713.

Wolday, F., Næss, P., Cao, X.J., 2019. Travel-based residential self-selection: A qualitatively improved understanding from Norway. Cities 87, 87–102. https://doi.org/10.1016/j.cities.2018.12.029.

Yang, S., Fan, Y., Deng, W., Cheng, L., 2017. Do built environment effects on travel behavior differ between household members? A case study of Nanjing, China. Transp. Policy 60, 1–17. https://doi.org/10.1016/j.tranpol.2017.12.006.

Yu, L., Xie, B., Chan, E.H.W., 2019. Exploring impacts of the built environment on transit travel: Distance, time and mode choice, for urban villages in Shenzhen, China. Transp. Res. Logist. Transp. Rev. 132, 57–71. https://doi.org/10.1016/j.tre.2019.11.004.

Zang, P., Lu, Y.i., Ma, J., Xie, B.o., Wang, R., Liu, Y.e., 2019. Disentangling residential self-selection from impacts of built environment characteristics on travel behaviors for older adults. Soc. Sci. Med. 238, 112515. https://doi.org/10.1016/j.socscimed.2019.112515.

Zhang, M., Zhang, W., 2020. When context meets Self-Selection: The built environment–travel connection revisited. J. Plan. Educ. Res. 40 (3), 304–319. https://doi.org/10.1177/0739456X18755495.

Zhang, Y., Zheng, S., Sun, C., Wang, R., 2017. Does subway proximity discourage automobility? Evidence from Beijing. Transp. Res. Part D Transp. Environ. 52, 506–517. https://doi.org/10.1016/j.trd.2016.11.009.

Zhao, P., Zhang, Y., 2018. Travel behaviour and life course: Examining changes in car use after residential relocation in Beijing. J. Transp. Geogr. 73, 41–53. https://doi.org/10.1016/j.jtrangeo.2018.10.003.