




ARTICLE



<https://doi.org/10.1057/s41599-024-02678-2>

OPEN

A new model for residential location choice using residential trajectory data

Yanzhe Cui¹, Pengjun Zhao ^{1,2✉}, Ling Li¹, Juan Li¹, Mingyuan Gong³, Yiling Deng⁴, Zihuang Si¹, Shuaichen Yan¹ & Xuewei Dang¹

Traditional residential location choice (RLC) models are based on the characteristics of location and demographics, revealing important patterns of RLC, but no RLC models have yet incorporated individual preferences. This study fills this gap by integrating the pattern of home-based travel into the RLC model. Firstly, by analysing residential trajectory data collected from Beijing and Shenzhen, we find that both residents' commuting time, that is, time spent commuting to work, and home-based non-commuting (HBNC) time, that is, time spent on the consumption of amenities when departing from homes, follow an extreme value distribution (EVT). This indicates that, based on time budget and financial constraints, residents strive to minimise commuting time and maximise HBNC time. Subsequently, by integrating these findings into individual-level RLC analysis, we obtain an RLC model that aligns with the gravity model. Throughout the model training process, we demonstrate that the RLC model exhibits strong robustness by incorporating control variables, changing the spatial scale of the observation unit, testing for endogeneity, and considering historical RLC. Moreover, the model performs well in applications including assessing dynamic changes in RLC behaviours and making predictions based on previous travel behaviours. The RLC model in this study advances our understanding of human habitat selection behaviour and can be utilised by policymakers to develop and implement effective urban planning and epidemic management policies.

¹School of Urban Planning and Design, Peking University Shenzhen Graduate School, Shenzhen 518055, China. ²College of Urban and Environmental Sciences, Peking University, Beijing 100871, China. ³School of Economics, Peking University, Beijing 100871, China. ⁴School of Design and Architecture, Zhejiang University of Technology, Hangzhou 310023, China. ✉email: pengjun.zhao@pku.edu.cn

Introduction

More than 55% of the world's population now live in cities (Vilar-Compte et al., 2021). Residential location choices (RLCs) have significant implications for the sustainable development of cities. From a city perspective, existing research shows that people's choices of residence can have a significant impact on the local economy (Li et al., 2013), spatial structure (De Vos et al., 2018; Næss et al., 2019), the environment (Engebretsen et al., 2018; Huu Phe and Wakely, 2000), the urban transport system (Taniguchi et al., 2014) and epidemic prevention and control (Liu and Tang, 2021). From an individual perspective, residential satisfaction can contribute significantly to overall life satisfaction (Campbell et al., 1976). Residential location modelling is therefore considered to be the core of one of the grand challenges of contemporary social science (Pagliara et al., 2010).

Numerous studies have been conducted to develop universal models of RLC in light of the importance of residential location. In general, existing RLC models can be categorised into two types. One type of research takes RLC models into account as an integral part of urban complex models (Alberverio et al., 2007; Baynes, 2009; Tonne et al., 2021), including MUSSA II, RELU-TRAN (Anas and Liu, 2007) and UrbanSim (Waddell, 2002). Models such as these are based on the interaction between the land market, labour market, the distribution of industry, and transportation to analyse RLCs (Ahlfeldt et al., 2015). Enabled by the use of massive data from multi-dimensions, they focus more on the interdependencies between sub-modules, rather than on the nature of location choices. The second type of RLC model explores factors that affect RLCs. Using the Multinomial Logit Model, they have examined the impact of individual characteristics and location characteristics on RLCs, such as age, gender, the number of family members and accessibility to infrastructure (Baum-Snow, 2007; Buzar et al., 2007; Campbell et al., 1976; Chen et al., 2016; Delgado and Bonnel, 2016; Garcia-López, 2012; Lee et al., 2010; Levinson, 2008; Melia et al., 2018; Portnov et al., 2011). Nevertheless, households and spaces can be characterised by a variety of dimensions, leading to unmanageable arrays or model specifications that are difficult to assemble for effective calibration (Pagliara et al., 2010).

Travel behaviours significantly impact RLCs (De Vos and Singleton, 2020), with studies indicating that people prefer to live in neighbourhoods that facilitate satisfying trips (De Vos and Witlox, 2016; Ettema and Nieuwenhuis, 2017). Low levels of travel satisfaction may encourage individuals to move to a different type of neighbourhood that allows for more frequent use of preferred modes of transportation (De Vos and Witlox, 2017). This illustrates that RLCs are shaped not just by amenities but also by personal preferences, for example, a car enthusiast may prefer to live in a suburban neighbourhood, or someone who enjoys walking or cycling may opt for an urban area (De Vos and Singleton, 2020). Therefore, the RLC model that focuses solely on the amenities fails to account for the influence of individual preferences on RLC. However, up to now, there has been no RLC model that takes individual travel behaviour into account. We aim to fill this gap by constructing an RLC model from the perspective of travel behaviours. During the modelling process, we rely on the allocation of travel time between home-based travels to build the RLC model, which not only diminishes the need for various types of data but also aids in simplifying the model's structure.

Big data and related analytics bring new opportunities for understanding RLCs. Human mobility data derived from spatiotemporal mobile phone trajectory data could be helpful to develop the travel-behaviour based RLC model. Mobile phone trajectory data has the advantage of a high sampling rate, large geographic coverage, low collection cost, and accurate information about space and time (Ni et al., 2018). Based on the time

budget and a working-resting timeframe, by combining mobile phone data with geocoded location information, we identify residential locations and workplaces through the comparison of stay durations across different times and places (Phithakkitnukoon et al., 2012; Yan et al., 2019; Zhao and Gao, 2023). Other locations where the stay exceeds 30 min are considered non-work sites. We can obtain a comprehensive picture of residents' home-based travels by analysing travel behaviour originating from or destined for residential locations, encompassing both commuting and non-commuting purposes. In this paper, we analyse trajectory data collected from over 16 million mobile phone users in three consecutive years between 2018 and 2020 in two megacities in China—Beijing and Shenzhen.

Compared to existing research, this paper has the following novelties: (1) We focus on analysing residents' revealed preferences rather than their stated preferences in RLCs. Revealed preferences are based on real decisions made in real-life situations. Stated preferences, on the other hand, are derived from what individuals say they would do, often in response to hypothetical scenarios, which may not always translate to actual behaviour due to biases or the hypothetical nature of the situation (Fujii and Gärling, 2003). Additionally, the use of mass mobile phone signalling data also reduces the issue of small samples, which is commonly encountered in stated preference studies (Thorhauge et al., 2016; Wan et al., 2021). (2) This study extends existing RLC models by considering individual residential preferences, which are proxied by home-based travel behaviours. We test the validity of the model in multiple ways, including adding control variables, changing the spatial scale of the observation unit, testing for endogeneity, and considering historical RLC. (3) This RLC model can be used not only to analyse the spatial distribution of residential locations at the group level, but also to analyse the RLC at the individual level. As an example of the model's application, we assess dynamic changes in RLC behaviours and make predictions based on previous travel behaviours.

Analytical framework

The RLC model, based on home-based travel behaviour, is developed, and Fig. 1 describes the process of our modelling. From the population's perspective, we construct the RLC model according to the gravity model (Batty et al., 1974), and from the viewpoint of individuals, we analyse RLCs based on the assumption of utility maximisation. Ultimately, the same RLC model is derived.

The gravity model and the RLC model. The population-level RLC model we employed is the constrained gravity model, which is essentially grounded on a balance between benefits and costs (Batty, 1983; Batty et al., 1974), as shown in Eq. (1).

$$T_{ij} = O_j P_{ij} = O_j \frac{m_i f(r_{ij})}{\sum_k m_k f(r_{ik})} \quad (1)$$

In this model, the attraction or benefit (m_i) of residing in any given location is weighed against the deterrence or cost ($f(r_{ij})$) to that location from another, with commuting commonly acknowledged as a form of deterrence (Barbosa et al., 2018; Pagliara et al., 2010). Owing to the constraints of financial budgets, the choice of residential location is inevitably affected by housing prices (DeSalvo and Huq, 1996; Zhuge et al., 2016). However, quantifying a region's attractiveness is quite challenging. Built environment and demographic characteristics are frequently seen as factors that affect a location's attractiveness (Bhat and Guo, 2007; Ettema and Nieuwenhuis, 2017; Schirmer et al., 2014), while relocation choice, which is also a form of RLC, is mainly influenced by individual

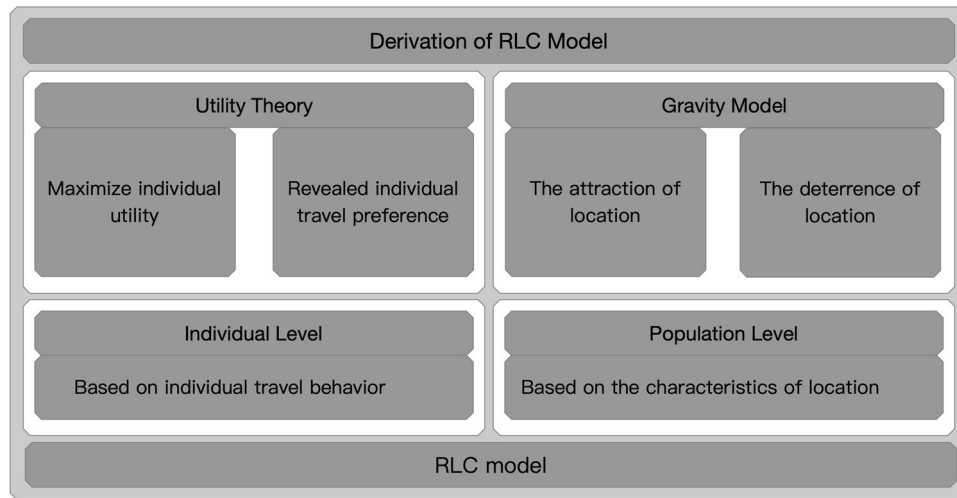


Fig. 1 Analytical framework.

preferences, such as the impact of historical factors and individual habits (Clark and Lisowski, 2017). However, to date, no studies have attempted to include individual preferences in the RLC model. In this study, we use HBNC time to represent a location’s attractiveness. Firstly, HBNC time is a reflection of residents’ revealed preferences, which can indicate their real needs. Secondly, HBNC travel is often for the consumption of built environment. The greater the demand for a certain amenity, the greater the weight of travel for these types of amenities in the HBNC time. Therefore, HBNC time includes information about an individual’s preferences. Compared to using amenities as a measure of a location’s attractiveness, using HBNC time as a proxy variable more closely aligns with our understanding because residents may choose a location mainly based on some of its built environment, rather than all of them. When

$$m_i = e^{\alpha \log(hc_i) + \gamma HBNC_time_{ij}} \tag{2}$$

$$f(r_{ij}) = e^{\beta C_time_{ij}} \tag{3}$$

where

$$C_time_{ij} = \frac{Time_{ij} + Time_{ji}}{N_{ij} + N_{ji}} \tag{4}$$

$$HBNC_time_{is} = \frac{Time_{is} + Time_{si}}{N_{is} + N_{si}} \tag{5}$$

where i is a residential location, j is a workplace and s is a non-work site, C_time_{ij} is the average commuting time for an individual in a month and $HBNC_time_{is}$ is his/her average HBNC time in the same month, $Time_{ij}$ ($Time_{ji}$) is the total travel time from residential location i (workplace j) to workplace j (residential location i), N_{ij} (N_{ji}) is the total number of trips from residential location i (workplace j) to workplace j (residential location i), $Time_{is}$ ($Time_{si}$) is the total travel time from residential location i (a non-work site s) to a non-work site s (residential location i) and N_{ij} (N_{ji}) is the total number of trips from residential location i (a non-work site s) to a non-work site s (residential location i).

Then, we can get the RLC model:

$$T_{ij} = O_j Prob_{ij} = O_j \frac{e^{\alpha \log(hc_i) + \gamma HBNC_time_{ij} + \beta C_time_{ij}}}{\sum_k e^{\alpha \log(hc_k) + \gamma HBNC_time_{kj} + \beta C_time_{kj}}} \tag{6}$$

where T_{ij} is the number of residents who work in location j and live in location i , O_j is the number of people who work in location j , $Prob_{ij}$ is the probability of residents choosing to live in location i

and work in location j , hc_i is the housing expenditure of location i , $HBNC_time_{ij}$ is home-based non-commuting time, C_time_{ij} is commuting time, and α , β and γ are parameters to be estimated. Consistent with the settings of quantitative spatial modelling, we include variables related to time in exponential form, while other variables are included in power-law form (Eaton et al., 2004; Hebllich et al., 2020).

The RLC model attempts to use residents’ travel behaviour and housing costs to explain jobs-housing relationship. The model’s dependent variable is the probability of a residential location being chosen, which means the model tries to figure out the distribution pattern of where the workforce lives in relation to their places of work. In comparison to traditional gravity models, our RLC model is not only simpler in form but also encompasses more information regarding individual preferences.

Utility maximisation and the RLC model. The individual-level RLC model is based on the assumption of utility maximisation (Ahlfeldt et al., 2015; Hebllich et al., 2020; Schirmer et al., 2014). We assume that the utility function of a risk-neutral resident o who works in location j and resides in location i is defined by the resident’s travel behaviour, housing expenditure and an idiosyncratic shock, as shown in Eq. (7). As commuting travel is a mandatory form of travel, we include commuting time (C_time_{ij}) in the utility function as an iceberg cost. HBNC time ($HBNC_time_{ij}$) has two parts, one that relates to consumption (αC_{ij}), and the other for travels (l_{ij}) that brings utility. Residents will make optimal choices regardless of individual preference differences, a heterogeneity parameter (z_{ijo}) which follows an extreme value distribution is thus included.

$$U_{ijo} = \frac{z_{ijo} w_j}{e^{\beta C_time_{ij}} hc_i} \left(\frac{\alpha C_{ij}}{\xi} \right)^\xi \left(\frac{l_{ij}}{1 - \xi} \right)^{1 - \xi} \tag{7}$$

$$s.t. \alpha C_{ij} + l_{ij} = e^{u HBNC_time_{ij}} \tag{8}$$

$$F(z_{ijo}) = e^{-z_{ijo}^{-\alpha}} \tag{9}$$

where w_j is the average wage level in location j . When individuals attempt to maximise U_{ijo} , the equilibrium utility is,

$$u_{ijo} = \frac{z_{ijo} w_j e^{u HBNC_time_{ij}}}{e^{\beta C_time_{ij}} hc_i} \tag{10}$$

By summing up the individual utilities, we can estimate the probability of choosing a residential location within a city. Hence,

the probability that a resident chooses to live in location i and work in location j is

$$\begin{aligned}
 Prob_{ij} &= \Pr[u_{ij0} \geq \max\{u_{rso}\}; \forall r, s] \\
 &= \frac{(e^{\beta C_{-timeij}} h_{c_i})^{-\alpha} (e^{\mu HBNC_{-timeij}} w_j)^{\alpha}}{\sum_{k=1} (e^{\beta C_{-timekj}} h_{c_k})^{-\alpha} (e^{\mu HBNC_{-timekj}} w_j)^{\alpha}} \quad (11) \\
 &\propto \frac{e^{\alpha \log(h_{c_i}) + \gamma HBNC_{-timeij} + \beta C_{-timeij}}}{\sum_k e^{\alpha \log(h_{c_k}) + \gamma HBNC_{-timekj} + \beta C_{-timekj}}}
 \end{aligned}$$

The $Prob_{ij}$ in individual-based RLC model follows the same structure as that in population-based gravity model. Although the form of the population-level model and the individual-level model is the same, the interpretation of the models differs: the former explains patterns in population spatial distribution, while the latter explains patterns in individual residence choices.

The generalisation and contribution of the RLC model. The generality of the RLC model in this study is reflected in the following two aspects: (1) The construction of the RLC model is based on both population-level method and individual-level method, providing a solid theoretical basis for examining the behaviour of both individuals and groups. (2) RLC analysis based on the gravity model has been applied in the West Midlands Conurbation in central England (Batty et al., 1974), while utility maximisation-based modelling analysis has been applied in London (Heblich et al., 2020) and Berlin (Ahlfeldt et al., 2015). These different applications illustrate the flexibility and effectiveness of the model’s base structure.

Our version of the RLC model adds a new dimension: residents’ preferences. The addition of this information adds greater depth to our understanding of how demographic factors impact where people choose to live. While our version of the RLC model introduces new analytical perspectives, variables and functional forms that differ from existing studies, our aim is to extend the application rather than to challenge previous RLC models.

Data and variables

Study area. We have selected two megalopolises in China as the area of study, namely Beijing and Shanghai. Beijing, the capital of China, maintained a stable population of 21 to 22 million from 2018 to 2020. It is situated in northern China, an inland city that does not border the sea. Shenzhen is situated in southern China and next to Hong Kong and had a population of 16.66 million in 2018, which has risen to 17.63 million in 2020 (statistics were drawn from China Statistical Yearbook). Both cities are economic hubs of their regions and have the highest GDP in their respective urban agglomerations. Based on statistics in 2018–2020, Beijing contributed 42% of the GDP in the Jing-Jin-Ji urban agglomeration, encompassing 13 cities; Shenzhen contributed over 30% of the GDP in the Pearl River Delta urban agglomeration, which includes 9 cities.

There are also significant differences between Beijing and Shenzhen. First, their geographical structures are different. Beijing is mostly situated on a plain, which allows for easy urban expansion, while Shenzhen’s expansion is restricted by hills and its coastline. According to the layout of residential locations, workplaces and home-based non-workplaces of both cities in Fig. 2, Beijing has a single centre, while Shenzhen shows a polycentric layout. Second, the two cities have different industrial structures. Beijing’s workforce is primarily engaged in IT, business services and finance, which require less industrial space. Shenzhen, on the other hand, has a significant manufacturing workforce (Chandra et al., 2023; Chen and Kenney, 2007). Third, administrative influences differ in the two cities. While Beijing, as the capital, is

Table 1 Distribution of people in categories for analysis.

Variable	Shenzhen		Beijing	
	Number of people	Percentage	Number of people	Percentage
Gender				
Female	1,422,528	32	4,475,160	36
Male	3,017,846	68	7,901,636	64
Age				
19–24	954,437	21	1,505,944	12
25–29	1,251,605	28	2,397,109	19
30–34	950,508	21	2,459,491	20
35–39	538,942	12	1,935,330	16
40–44	299,452	7	1,303,226	11
45–49	230,966	5	1,245,858	10
50–54	145,921	3	919,514	7
55–59	68,543	2	610,324	5
Total	4,440,374	100	12,376,796	100

subject to more top-down government decisions regarding urban planning, Shenzhen, as a special economic zone, has fewer administrative restrictions.

Datasets and data processing

Mobile signalling data. We test the above RLC model with spatiotemporal travel trajectory data extracted from more than 4 million regular mobile phone users in Shenzhen and more than 12 million regular mobile phone users in Beijing (see Table 1). The main data is mobile phone signalling data, with trajectories derived from the time the user communicated with a base station and the coordinates of the base station. We selected samples from November 2018, November 2019 and November 2020, specifically choosing those that appeared more than 10 days in a month. To reduce the impact of extreme values, commuting time over 180 min and HBNC time over 300 min were excluded. Due to COVID-19 starting in early January 2020, our pre-pandemic months include November 2018 and November 2019, while the post-pandemic period includes November 2020, allowing us to test the effectiveness of the RLC model following the pandemic.

The individual’s coordinate point position was calculated by the Operator using a multi-base station weighting algorithm. According to the Operator’s processing logic, points with a stay of more than 30 min are considered stay points. Moreover, the workplace is the longest stay point during the weekdays from 5 a.m. to 8 p.m., and the residence is the longest stay point from 8 p.m. to 5 a.m. Using these details, along with the start stay point and the end stay point for each trip and their exact time, we calculated the duration of each trip, namely travel time, identified the purpose of each trip and counted the number of each type of trips. Based on the analysis mentioned above, we can obtain the residents’ commuting time and HBNC time (see Table 2). Due to the Operator’s data protection rules, we can only extract the values of the above variables in a squared grid or tiles. Notably, only tiles with more than 5 identified residents were considered. To process the data, the study areas was divided into squared tiles, and we took the monthly average of commuting time and HBNC time of residents with residences falling in the same tile.

Housing expenditure. The housing data we used include housing prices and government guideline prices. Housing prices refer to the listed prices of individual housing units, which were obtained from public websites. We have provided statistical descriptions of our housing price data in Table 2. However, there is an issue that in some areas, the number of housing units listed may be limited,

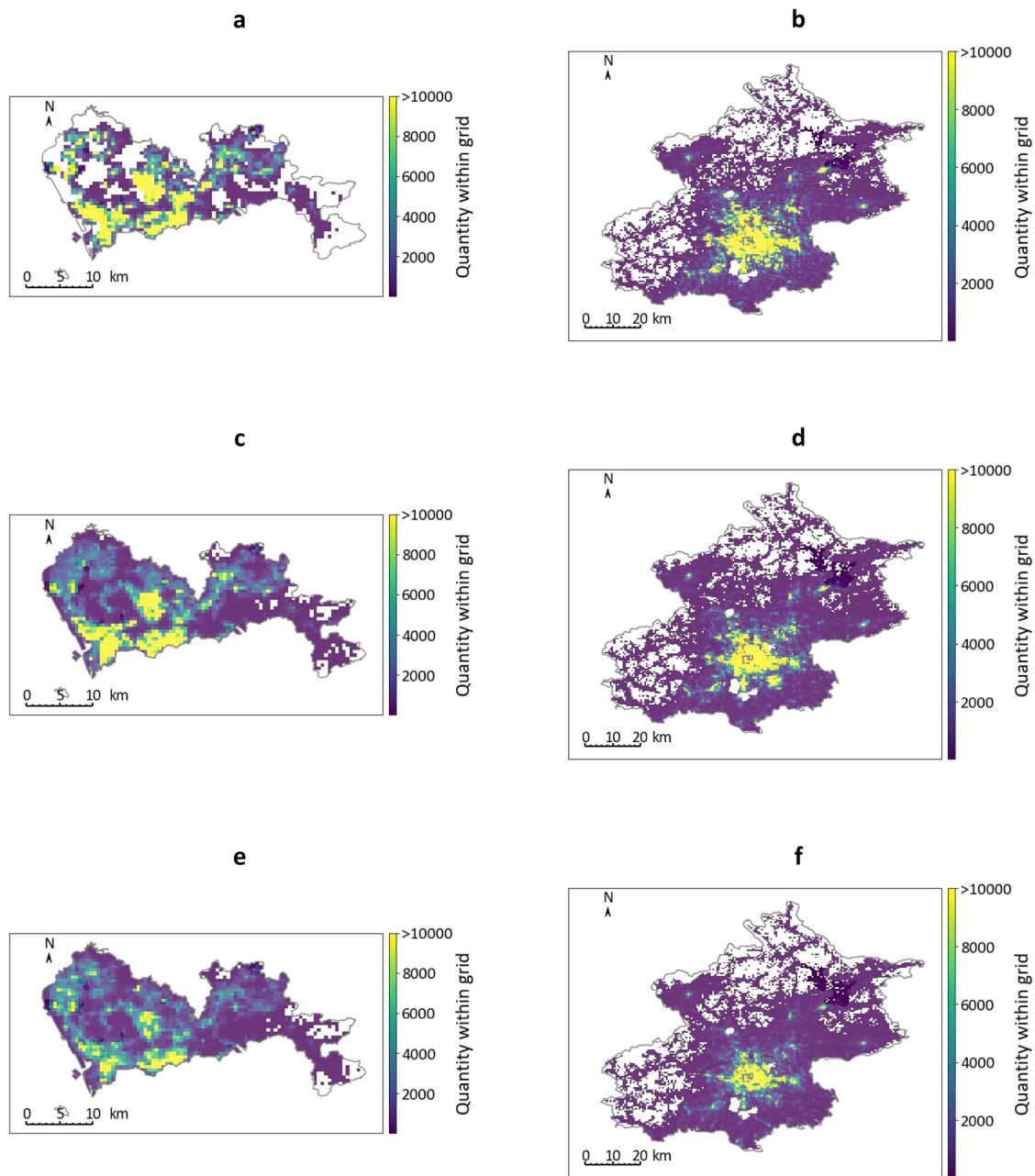


Fig. 2 The distribution of workplaces, residential locations and home-based non-workplaces in Shenzhen and Beijing. Graphs (a) and (b) depict the distribution of residential locations in Shenzhen and Beijing, respectively. Graphs (c) and (d) depict the distribution of workplaces in Shenzhen and Beijing, respectively. Graphs (e) and (f) depict the distribution of Home-based non-workplaces in Shenzhen and Beijing, respectively.

leading to an inaccurate representation of the area. To minimise this error, we calculated the average listing price for each neighbourhood (referred to as ‘jiedao’, the smallest administrative unit within a city) and then assigned this average price to each tile based on the jiedao where the centre of the tile is located.

Other data. We also utilised Point of Interest (POI) data, which are all publicly accessible from OpenStreetMap. These data were associated with each tile to generate control variables for the RLC model. This primarily included calculating the distance from the centre of each tile to the nearest subway, bus stations, hospitals, retail markets, parks and schools (Næss, 2006a; Næss et al., 2019; Rivas et al., 2019; Sander, 2006). To validate the robustness of the RLC model using the instrumental variables method, we also used precipitation data.

Empirical implementations

Our empirical analysis consists of two parts: model verification and model application, as shown in Fig. 3. The individual-level RLC model posits that individuals’ idiosyncratic preferences, which adhere to the Extreme Value Theory (EVT), are crucial. Therefore, we employ a fitting analysis method to determine whether residents’ travel behaviour aligns with an extreme distribution. Next, the RLC model is fitted using Generalised Linear Models (GLM) and verified by adding control variables, using instrumental variables and analysing the impact of scale effects (Barbosa et al., 2018). Finally, we utilise the RLC model to examine shifts in residential location preferences due to COVID-19 and to assess whether it can accurately capture dynamic changes in RLC, as well as to make forecasts based on historical travel patterns.

Table 2 Statistical descriptions of commuting time (minutes), HBNC time (minutes) and housing prices (yuan).

		Shenzhen			Beijing		
		2018	2019	2020	2018	2019	2020
commuting time	Number of obs.	1,015,403	1,130,783	1,064,446	4,362,023	4,458,744	4,253,309
	mean	20.21	24.82	23.37	33.45	31.61	32.68
	std	20.78	22.20	20.83	27.44	26.61	27.20
	min	0	0	0	0	0	0
	25%	6.06	8.03	7.35	12.40	10.83	11.59
	50%	13.98	18.49	17.60	26.90	25.22	26.16
	75%	27.20	35.41	33.61	46.97	45.10	46.35
	max	180	180	180	180	180	180
HBNC time	Number of obs.	1,015,403	1,130,783	1,064,446	4,362,023	4,458,744	4,253,309
	mean	28.55	24.58	23.82	28.90	26.13	28.25
	std	29.21	19.83	18.83	21.73	18.75	21.86
	min	0	0	0	0	0	0
	25%	10.91	12.89	12.20	15.58	14.45	14.77
	50%	20.24	20.38	20.05	24.09	22.26	23.22
	75%	35.86	30.69	30.53	36.12	32.98	35.50
	max	300	300	300	300	300	300
housing prices	Number of obs.	17,068	20,144	10,961	11,780	12,305	11,806
	mean	63234.43	57334.81	55359.03	62659.74	62667.48	63323.23
	std	24591.32	20418.53	17647.98	36804.24	38871.89	37171.76
	min	12014.0	10049.0	10160.0	10047.0	10056.0	10056.0
	25%	44097.5	42239.25	42532.0	34549.75	33059.0	35000.0
	50%	57862.0	54080.0	52958.0	53669.5	53510.0	54223.0
	75%	78379.5	68198.25	64584.0	85071.25	86263.0	85989.5
	max	213513.0	171945.0	149477.0	266976.0	278100.0	270000.0

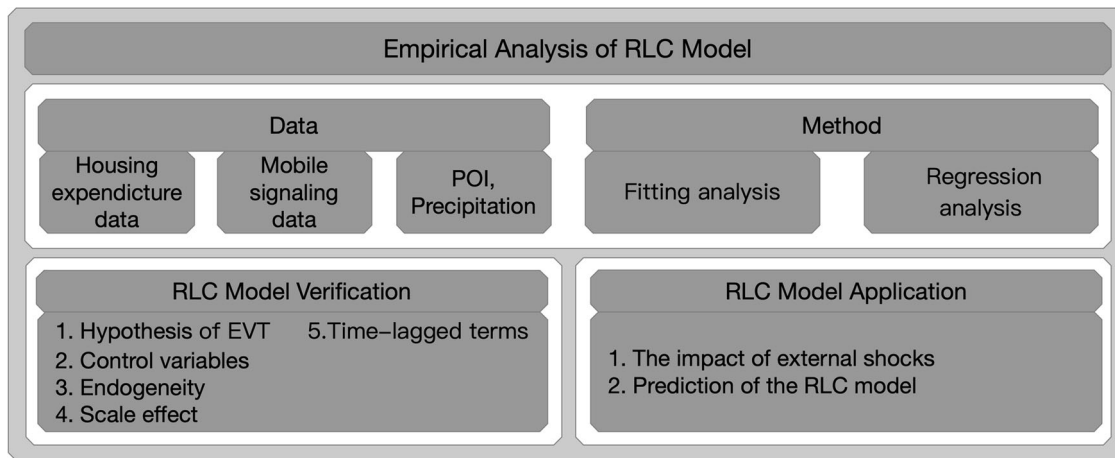


Fig. 3 Road map of empirical analysis.

Verification of the RLC model

Home-based travel behaviour and EVT. Individuals’ idiosyncratic preferences aligning with the EVT is a crucial hypothesis in our RLC model. Given that each tile may contain a different number of people, we assign a weight to each tile based on the number of included residents. We then use the Generalized Extreme Value (GEV) distribution to check if commuting time and HBNC time align with EVT. The fitting results for commuting time in Fig. 4 show that residents selected a residential location that enables them to achieve minimum commuting time, given the spatial distribution of amenities and housing prices. Likewise, the fitting results for HBNC time in Fig. 4 show that HBNC time is maximised during RLC. This means that the way people travel from home aligns with our model’s hypothesis.

There are reasonable explanations for the above findings. Travel is primarily driven by the expected benefits at the

destination (Næss et al., 2019; Wang et al., 2018). While travel time constitutes a cost paid to participate in out-of-home activities, its impact on individual utility is highly dependent on whether activities are mandatory or optional (Ye et al., 2020). Commuting is rigid travel since work is the primary source of income, and stress-related effects (high blood pressure, self-reported tension and reduced task performance) may extend beyond the journey itself (Kluger, 1998). As a result, it is seen as unproductive time (Lyons and Chatterjee, 2008). Comparatively, HBNC travel offers greater flexibility, since residents not only have the option of choosing the departure time and destination of their trips, but also whether to travel. In other words, residents can decide not to travel to a particular destination if the travel cost is greater than the utility gained at that location. By maximising HBNC trips derived from leisure time, residents can increase their utility. In comparison to distance indicators between residential locations and amenities (schools, parks, etc.)

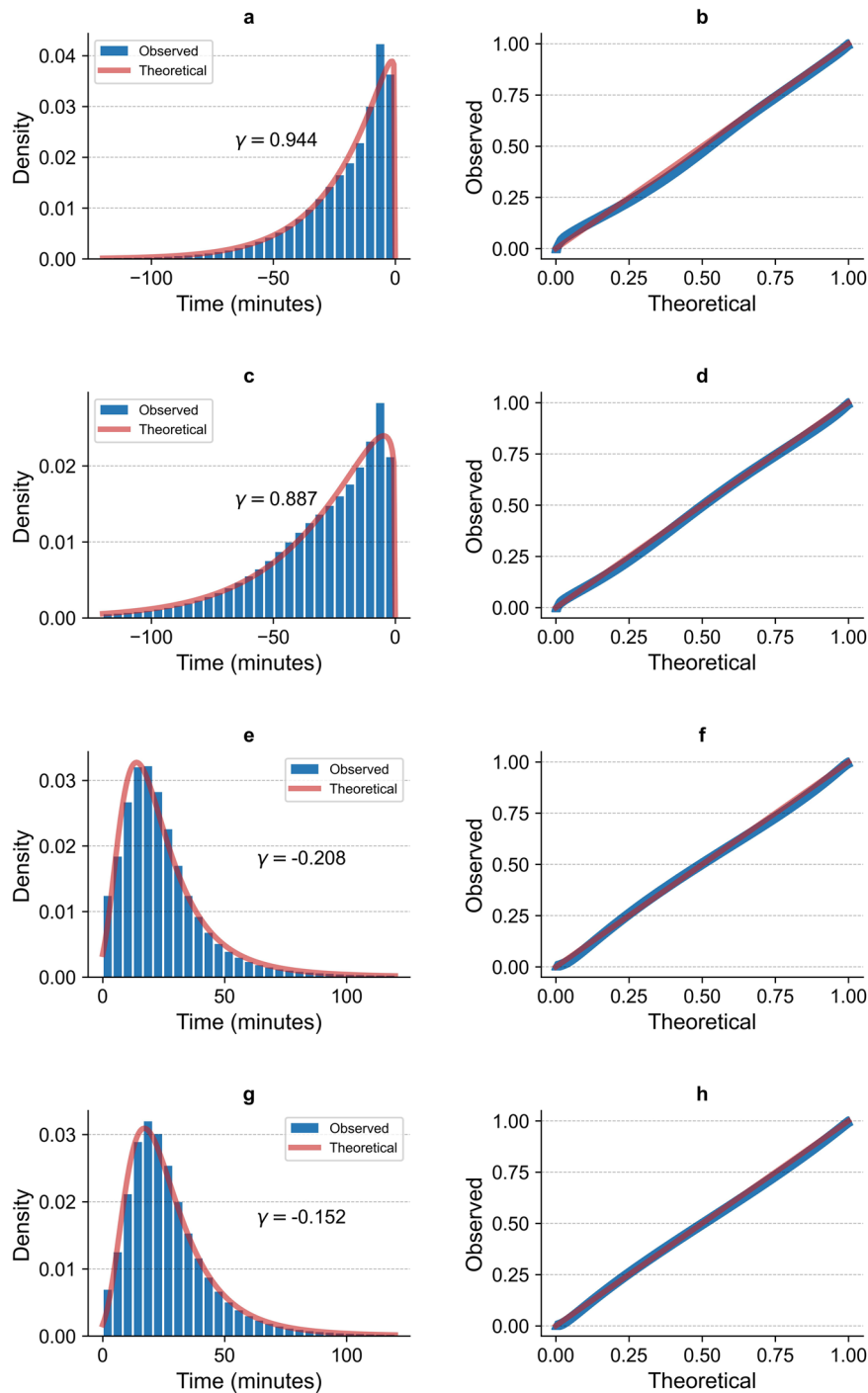


Fig. 4 Sample distribution of home-based travel time and corresponding fitted GEV distribution. Graphs (a–d) are the sample distribution of commuting time and corresponding fitted GEV distribution. Graphs (a, b) are the fitting diagram and P-P plot for Shenzhen, respectively. Graphs (c, d) are the fitting diagram and P-P plot for Beijing, respectively. Graphs (e–h) are the sample distribution of HBNC time and corresponding fitted GEV distribution. Graphs (e, f) are the fitting diagram and P-P plot for Shenzhen, respectively. Graphs (g, h) are the fitting diagram and P-P plot for Beijing, respectively.

which are primarily a reflection of the accessibility of amenities, HBNC travel reflects people’s personal preferences as well.

According to the fitting results for commuting time and HBNC time, we find that the concentration degree of commuting time and HBNC time for Beijing residents is higher than that for Shenzhen residents. The reason for this phenomenon is possibly due to the differences of the two cities in urban structure and natural characteristics. Beijing is a single-centre city (Yang et al., 2021), and urban expansion is not limited by space. In contrast,

Shenzhen is a polycentric city (Lai et al., 2022), where mountains, rivers and seas largely constrain the city’s expansion.

Regression analysis of the RLC model. In this section, we first analyse whether the results of the RLC model conform to our expectation, and then discuss the robustness of the results. Fitted using GLMs, a consistent pattern of parameters is observed in both Beijing and Shenzhen, despite the differences in their spatial structures (Table 3). The regression results of RLC model show

Table 3 Regression results of RLC model.

	Shenzhen		Beijing	
Commuting time	-0.0316*** (-35.8296)	-0.0315*** (-35.7446)	-0.0137*** (-33.8961)	-0.0138*** (-34.3468)
HBNC time	0.0058*** (6.519)	0.0059*** (6.6898)	0.005*** (9.4249)	0.0046*** (8.5995)
Housing price	-1.7164*** (-46.2985)	-1.6303*** (-41.8073)	-0.7952*** (-48.6038)	-0.6974*** (-40.0963)
Subway		0.0005 (0.3505)		0.0119*** (8.6192)
Bus station		0.0047*** (5.3739)		0.0045*** (8.9127)
Hospital		0.0021* (2.4035)		0.0083*** (9.071)
Retail market		-0.0004 (-0.4039)		0.0072*** (7.1413)
Park		0.001 (0.8562)		0.0029** (3.0377)
School		0.0054*** (6.1903)		0.0021* (2.3358)
Constant	16.2156*** (39.834)	15.334*** (35.8264)	5.5601*** (32.53)	4.421*** (23.8589)
R squared	0.1592	0.1631	0.0581	0.0651
Observations	21221	21221	56516	56516

*, **, and *** represent significance at the 5%, 1% and 0.1% levels, respectively.

Table 4 Wald test on coefficients of key variables.

	Beijing	Shenzhen
Coefficient of Commuting time -	0.0087***	0.0258***
Coefficient of HBNC time		
Control	No	No
Wald χ^2	675.2508	702.3446
$P > \chi^2$	0.0000	0.0000

Coefficients are from Eq. (6). *** represents significance at the 0.1% level.

that the probability of a tile being chosen as residential location decreases as the average commuting time within the tile increases (Commuting time was significantly negatively correlated with $Prob_{ij}$) and the probability of the tile being chosen as residential location increases as the average HBNC time within the tile increases (HBNC time was significantly positively associated with). In addition, the housing expenditure in a given tile was inversely related to the probability of that tile being selected as residential location. That is, $\alpha, \beta < 0$ and $\gamma > 0$, which is consistent with our expectations.

The more mandatory the activity, the greater the influence on the location choice of residence (As, 1978; Stopher et al., 1996). Hence we compare the coefficients of HBNC time and commuting time using the Wald test. The results in Table 4 show that the coefficient size of commuting time is significantly larger than that of HBNC time, suggesting a greater impact of commuting time on RLCs. As compared to HBNC travel, commuting travel is more mandatory. The destinations for HBNC travel are, in most cases, highly substitutable, while the workplace is generally more rigid. In addition, commuting travel is a prerequisite for HBNC travel, especially maintenance travel related to consumption, such as grocery shopping and medical appointments (Loa et al., 2021). Therefore, this result is in line with our expectations.

Robustness test 1: control variables. Amenities have an impact on RLCs (Campbell et al., 1976). To reduce errors caused by omitted variables, amenity variables are added to the RLC model to test the impact of missing variables. The results in Table 3 indicate that there were no significant changes in the significance and sign of the core explanatory variables, demonstrating the robustness of our RLC model. The HBNC time proposed in this study not only reflects the convenience of amenities associated with the residence but also the residents' revealed preferences. Therefore, HBNC time can, to a certain extent, act as a proxy for

these amenities. We observed changes in the explanatory power of the model by adding control variables to it. As shown in Fig. 5 (see Table 3 and Supplementary Tables 1–2), including control variables improved the model's goodness of fit (i.e., R^2) by 2% in Shenzhen and by 11% in Beijing. Similar modest changes are noted in the coefficient of HBNC time, especially in Shenzhen, but the change in the coefficient of commuting time is negligible in both cities. Hence, HBNC time serves as a good proxy for the availability of amenities and individual preferences for these amenities in both Shenzhen and Beijing.

Robustness test 2: endogeneity. Although the results of the model are significant, there may still be self-selection bias. For example, aggregation will promote the increase of infrastructure, and the increase of infrastructure will lead to further aggregation. We use instrumental variable framework to verify the robustness of the RLC model. As our RLC model is based on human mobility, weather is an ideal instrument (Aral and Nicolaides, 2017). Gender and age could cause gaps in commuting, income and individual preferences (Dökmeci and Berköz, 2000; Fuchs, 1986; Green and Hendershott, 1996; Huebner and Pleggenkuhle, 2015; Shin and Tilahun, 2022; Venter et al., 2007). We used the amount of precipitation per month per tile, the percentage of age per tile, and the percentage of gender per tile as instrumental variables, employing two-stage least squares method for the examination of endogeneity (see Supplementary Table 3). All groups passed the weak identification test, indicating that our model is robust.

Robustness test 3: scale effect. Due to the use of mobile signalling data in this study, the accuracy of individual positions will increase with the size of the tile. Therefore, we need to test the robustness of the RLC model on different scales. The platform developed by the operator provides tiles of 250 m × 250 m. Based on this, we further divide the two cities into tiles of 500 m × 500 m, tiles of 1000 m × 1000 m and tiles of 2000 m × 2000 m, respectively. Through training our RLC model at different scales, we find that housing prices, commuting time and HBNC time all register consistent coefficients that are significant at the 1% level, despite modest changes in the coefficient size (see Supplementary Tables 4–9). This indicates that our model is applicable at different scales.

The relative importance of commuting time and HBNC time is also examined at different spatial scales. To assess the impact of these two factors, a new index(RAV) is created. As shown in Eq. (12), this index is the absolute value of the ratio of the commuting

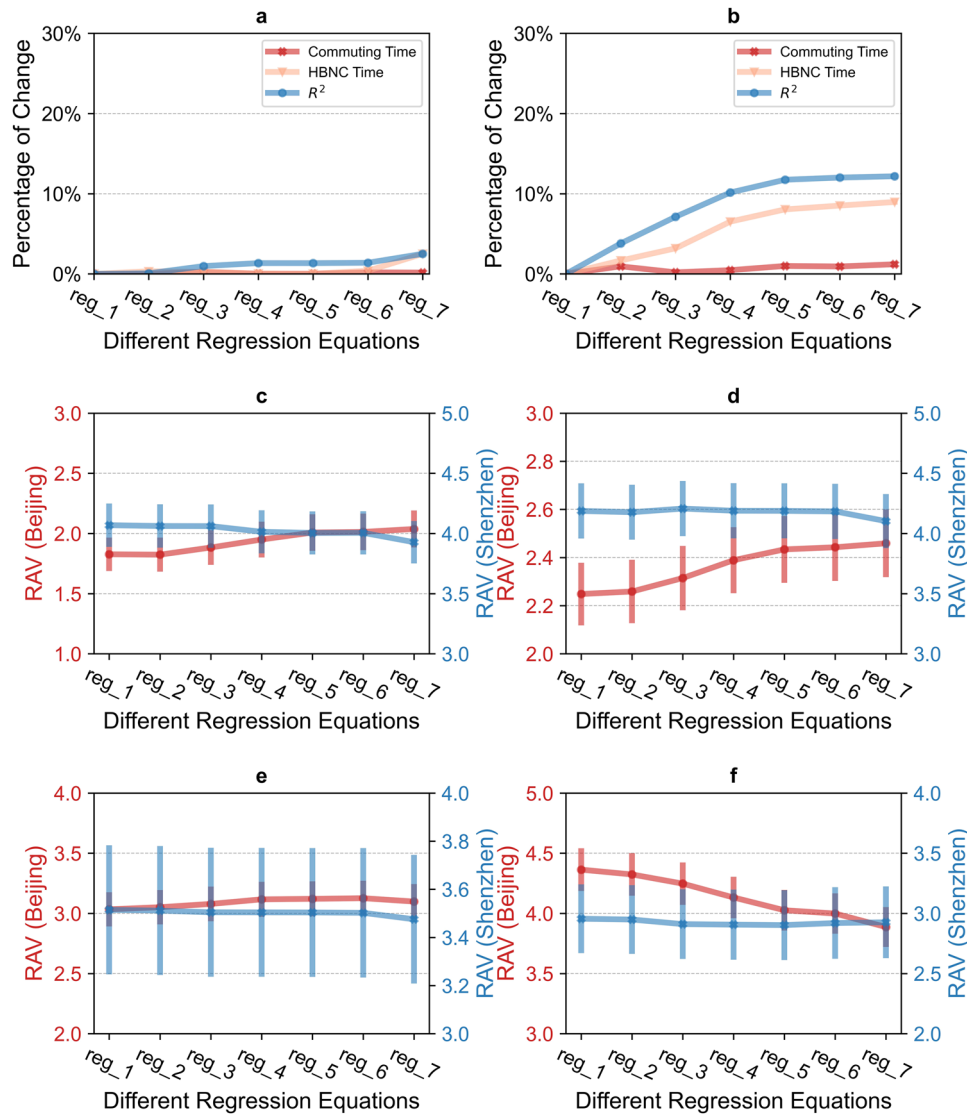


Fig. 5 The results of robustness test. Graphs (a, b) depict the percentage change in coefficient and goodness of fit with the inclusion of amenity variables in Shenzhen and Beijing, respectively. The vertical axis indicates the percentage change of the coefficient of HBNC time/coefficient of commuting time/goodness of fit of the regression model. On the horizontal axis, reg_1 is the baseline regression, and reg_2 to reg_7 represent regressions that gradually add the variables of subway, bus station, hospital, retail market, park and school. Graphs (c–f) show the relative importance of commuting time and HBNC time in determining residential locations on diverse spatial scales (tiles of 250 m × 250 m, 500 m × 500 m, 1000 m × 1000 m, and 2000 m × 2000 m, respectively). To ensure the reliability of the results, we choose the maximum value of the coefficients of commuting and HBNC time at the 5% confidence level.

time coefficient to the HBNC time coefficient.

$$RAV = \left| \frac{\text{Coefficient of commuting time}}{\text{Coefficient of HBNC time}} \right| \quad (12)$$

RAV greater than 1 indicates that commuting time has a greater impact than HBNC time. As shown in Fig. 5, commuting time has a consistently greater impact on the choice of residential location across different scales compared to home-based non-commuting (HBNC) time. This is in line with our expectations, therefore, we consider the results to be robust.

Robustness test 4: time-lagged terms. We incorporate time-lagged term in the model to test its robustness, which is inspired by prospect theory and the collective mobility model (Clark and Lisowski, 2017; Xu et al., 2021). When other conditions remain constant, it is possible to explain current RLCs by using historical RLCs. After including the probability of a residential location

being chosen in the previous period, as shown in Table 5, all results are consistent with the baseline regression. The goodness of fit of the models in both cities has improved significantly, suggesting that the choice of current location is significantly influenced by historical residential location distribution.

Application of the RLC model. We explore two applications of our RLC model. First, whether external shocks will affect the applicability of the RLC model. As a result of an exogenous disruption that eliminates the cues that trigger individual behaviours, people are forced to resort to deliberate decision-making (Verplanken et al., 2008; Verplanken and Wood, 2006) and make rational changes regarding their residential locations. Considering that rational choice is a fundamental assumption in our modelling process, we expect the RLC model to capture such changes. Second, to what extent our RLC model can be used for predictions. Prior research has confirmed the predictive power of RLC

Table 5 Regression results for the RLC model to capture the change in jobs-housing relationship.

	Shenzhen		Beijing	
Commuting time	-0.0373*** (-37.1862)	-0.0374*** (-37.1651)	-0.0187*** (-38.0311)	-0.0187*** (-37.9709)
HBNC time	0.0152*** (13.0772)	0.0152*** (13.0622)	0.0046*** (7.2259)	0.0043*** (6.7304)
Housing price	-1.4203*** (-33.9717)	-1.3961*** (-31.8619)	-0.7049*** (-36.8633)	-0.6322*** (-31.1082)
Prob.t-1	0.2824*** (39.2683)	0.2825*** (38.7239)	0.2854*** (55.458)	0.2803*** (54.2488)
Control variables	No	Yes	No	Yes
Constant	13.0797*** (28.5988)	12.7433*** (26.5667)	4.8769*** (24.4585)	4.0099*** (18.5464)
R squared	0.2965	0.2989	0.1469	0.1508
Observations	14410	14410	38219	38219

*** represents significance at the 0.1% level.

models based on amenities and population characteristics. Our model, which focuses on travel behaviour, not only considers spatial characteristics but also individual travel preferences. We therefore expect good predictive power of our RLC model.

The impact of external shocks. We consider COVID-19 as an external shock and test its impact on RLCs through our model. To minimise the risk of infection, many individuals have begun to work and study remotely, as well as reducing their optional travel after the breakout of COVID 19 (Zhang et al., 2021). There are concerns that the pandemic may have changed residents' living and working patterns (Gerwe, 2021; Liu and Tang, 2021). Therefore, we estimate the parameters of the RLC model separately for the pre-pandemic and post-pandemic periods to test the impact of the pandemic. According to the results (see Supplementary Tables 10–25), neither the sign nor the significance of commuting time or HBNC time has changed following the pandemic. *RAV* remains larger than 1, indicating that the relative importance between commute and non-commute travel has not changed. However, we observe a significant increase in the *RAV*, as shown in Fig. 6. This indicates that, as a result of the pandemic, commuting time has become more influential on residential location decisions than HBNC time. Due to safety considerations, each trip not only requires thinking about the utility it brings but also the risk of infection. Thus, the importance of HBNC time in the decision-making process diminishes, for instance, residents have noticeably reduced their use of amenities (Yu et al., 2023).

Prediction of the RLC model. In this section, we assess the predictive power of the RLC model. Initially, we utilise 2019 data to train the model, which is then employed to forecast individuals' RLCs for 2020. The predictive power of the model is assessed by contrasting the actual and forecasted values for 2020, as depicted in Fig. 7. It is noticeable that the model's predicted values have a positive correlation with the actual values across various spatial scales. Furthermore, since we are using only a subset of the urban population in our sample, to reduce the errors brought by magnitude, we draw on the methods of ordinal utility theory and compare the differences between the predicted ranks and the actual ranks, as shown in Fig. 7. It is evident that the predicted ranks from the model also show a positive correlation with the observed ranks across all spatial scales.

Conclusions

In a rapidly expanding urban environment, residents are experiencing both the convenience of agglomeration and its negative externalities (Arnott, 2007; Hong et al., 2020; Peng et al., 2017). RLC is essential not only for residents' life satisfaction (Campbell et al., 1976), but also for the urban spatial structure

(Næss, 2006b). Exploring RLC patterns is therefore a critical global issue. In this context, we develop an RLC model based on home-based travel and housing expenditure. This model aligns with both the population-level gravity model and the individual-level utility maximisation model. Analysing trajectory records of over 16 million mobile phone users from Beijing and Shenzhen across three years, we ascertain two main points: (1) residents aim to minimise commuting time, aligning with existing research (Guidon et al., 2019; Jang and Yi, 2021), and (2) they seek to maximise HBNC time. The RLC model is not only robust but also demonstrates broad applicability: (1) it suits cities with varying urban structures and geographical features, (2) it is valid across different spatial scales and regressions, (3) it can detect the effect of external shock and be used for prediction.

This paper offers a novel perspective on analysing RLC behaviour, not only incorporating individual preferences into the RLC model but also reducing data demands and diminishing the statistical correlation between sub-modules of the urban complex model (Anas and Liu, 2007; Waddell, 2002). The model is capable of explaining patterns of residence choice, as well as forecasting housing demand because of its strong predictive performance. Furthermore, since our RLC model is based on revealed preferences, the model can be combined with other models based on spatial characteristics to evaluate the efficiency of infrastructure provision and the impact of external shocks on the jobs-housing relationship (Næss, 2006b; Næss et al., 2019).

Although the proposed RLC model has many advantages, there are several limitations that need to be mentioned. First and foremost, there may be omitted variables. Our RLC model is constructed based on residents' travel behaviours, and it has included factors related to the built environment that are associated with travel. Nevertheless, it does not consider factors such as the noise and air quality, which can influence RLCs but are less related to travel behaviour. In future studies, these environment variables should be better considered. Secondly, we obtained secondary travel trajectory data rather than original call detail records. There is no way to verify the quality of the travel behaviour data which is essential to the test of our RLC model. Although the same dataset has been applied in published works, there remains the need to cross-check its reliability. While the use of individual travel trajectory data is limited due to data security concerns, this affects the accuracy of our analysis in the empirical tests of our RLC model. Moreover, a binary distinction is made between mandatory and optional travels, thereby reducing the accuracy of using HBNC time as a proxy for amenities and individual preferences. In addition, HBNC time was underestimated because of the ignorance of co-occurrences of non-work site visits. That is, we failed to account for leisure travels made outside homes. Future studies may attempt to justify the laws found in this paper by identifying the different types of HBNC

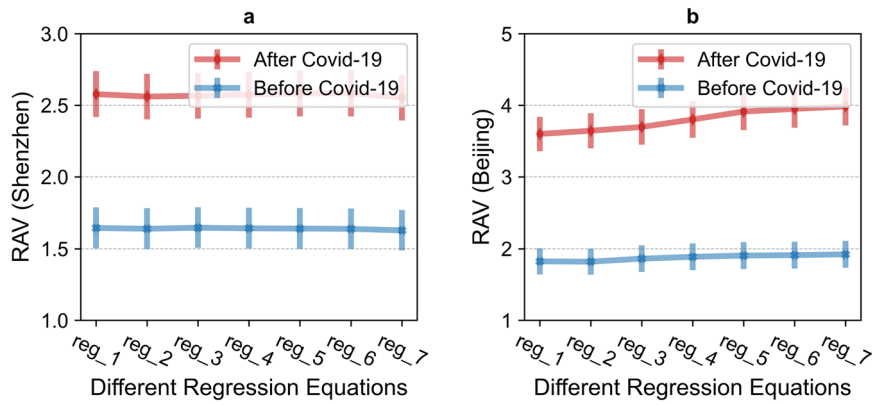


Fig. 6 The absolute value of the ratio between the coefficient of commuting time and the coefficient of HBNC time before and after the pandemic. Graphs (a, b) test the change in the impact of home-based travel time due to the outbreak of COVID-19 in Shenzhen and Beijing, respectively. Reg_1 to reg_7 correspond to those used in Fig. 5.

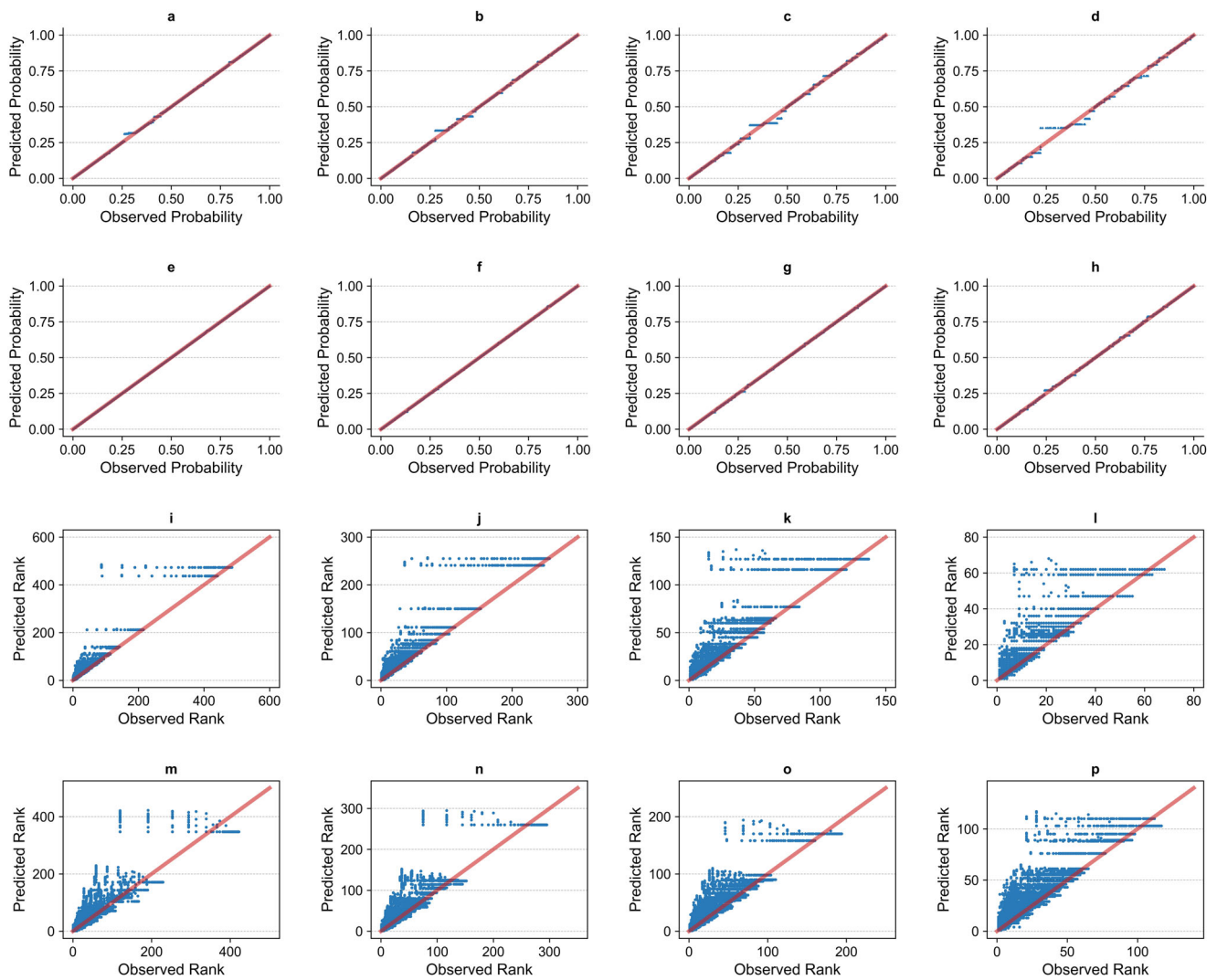


Fig. 7 The prediction results of RLC model. Graphs (a–h) are the P-P plot for predicted residential location distribution and observed residential location distribution. Graphs (a–d) are the P-P plot for Shenzhen on grid with tiles of 250 m × 250 m, 500 m × 500 m, 1000 m × 1000 m and 2000 m × 2000 m, respectively; Graphs (e–h) are the P-P plot for Beijing on the grid with tiles of 250 m × 250 m, 500 m × 500 m, 1000 m × 1000 m and 2000 m × 2000 m, respectively. Graphs (i–p) are the scatter for predicted residential location rank and observed residential location rank. Graphs (i–l) are the scatter for Shenzhen on the grid with tiles of 250 m × 250 m, 500 m × 500 m, 1000 m × 1000 m, and 2000 m × 2000 m, respectively; Graphs (m–p) are the scatter for Beijing on grid with tiles of 250 m × 250 m, 500 m × 500 m, 1000 m × 1000 m and 2000 m × 2000 m, respectively.

travels, which will help improve the explanatory power and predictive ability of this model.

Data availability

The data that support the findings of this study are available from China United Network Communications Group Co., Ltd., but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data is available upon reasonable request and with permission of China United Network Communications Group Co., Ltd.; Precipitation data can be accessed at <https://doi.org/10.5281/zenodo.3185722>; Government guideline prices data can be found in <https://zjj.sz.gov.cn/attachment/0/749/749839/8545737.pdf>; Housing prices data can be found in <https://cm.lianjia.com/>; Point of Interest (POI) data can be found in <https://www.openstreetmap.org>.

Received: 14 September 2023; Accepted: 12 January 2024;

Published online: 12 February 2024

References

- Ahlfeldt GM, Redding SJ, Sturm DM, Wolf N (2015) The economics of density: evidence from the Berlin Wall. *Econometrica* 83(6):2127–2189
- Alberverio S, Andrey D, Giordano P, Vancheri A (2007) The dynamics of complex urban systems: an interdisciplinary approach. Springer
- Anas A, Liu Y (2007) A regional economy, land use, and transportation model (relu-tran©): formulation, algorithm design, and testing. *J Regional Sci* 47(3):415–455
- Aral S, Nicolaides C (2017) Exercise contagion in a global social network. *Nat Commun* 8(1):14753
- Arnott R (2007) Congestion tolling with agglomeration externalities. *J Urban Econ* 62(2):187–203
- As D (1978) Studies of time-use: problems and prospects. *Acta Sociol* 21(2):125–141
- Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: Models and applications. *Phys Rep* 734:1–74
- Batty M (1983) A strategy for generating and testing models of migration and urban growth. *Regional Stud* 17(4):223–236
- Batty M, Hall P, Starkie D (1974) The impact of fares-free public transport upon urban residential location. *Proc Transport Res Forum* 15(1):347–353
- Baum-Snow N (2007) Did highways cause suburbanization? *Q J Econ* 122(2):775–805
- Baynes TM (2009) Complexity in urban development and management: Historical overview and opportunities. *J Ind Ecol* 13(2):214–227
- Bhat CR, Guo JY (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transport Res Part B: Methodol* 41(5):506–526
- Buzar S, Ogden P, Hall R, Haase A, Kabisch S, Steinführer A (2007) Splintering urban populations: emergent landscapes of reurbanisation in four European cities. *Urban Stud* 44(4):651–677
- Campbell A, Converse PE, Rodgers WL (1976) *The quality of American life: perceptions, evaluations, and satisfactions*. Russell Sage Foundation
- Chandra K, Wang J, Luo N, Wu X (2023) Asymmetry in the distribution of benefits of cross-border regional innovation systems: the case of the Hong Kong-Shenzhen innovation system. *Regional Stud* 57(7):1303–1317
- Chen K, Kenney M (2007) Universities/research institutes and regional innovation systems: the cases of Beijing and Shenzhen. *World Dev* 35(6):1056–1074
- Chen Y, Lü B, Chen R (2016) Evaluating the life satisfaction of peasants in concentrated residential areas of Nanjing, China: a fuzzy approach. *Habitat Int* 53:556–568
- Clark WA, Lisowski W (2017) Prospect theory and the decision to move or stay. *Proc Natl Acad Sci* 114(36):E7432–E7440
- De Vos J, Ettema D, Witlox F (2018) Changing travel behaviour and attitudes following a residential relocation. *J Transp Geogr* 73:131–147
- De Vos J, Singleton PA (2020) Travel and cognitive dissonance. *Transport Res Part A: Policy Pract* 138:525–536
- De Vos J, Witlox F (2016) Do people live in urban neighbourhoods because they do not like to travel? Analysing an alternative residential self-selection hypothesis. *Travel Behav Soc* 4:29–39
- De Vos J, Witlox F (2017) Travel satisfaction revisited. On the pivotal role of travel satisfaction in conceptualising a travel behaviour process. *Transport Res part A: policy Pract* 106:364–373
- Delgado JC, Bonnel P (2016) Level of aggregation of zoning and temporal transferability of the gravity distribution model: the case of Lyon. *J Transp Geogr* 51:17–26
- DeSalvo JS, Huq M (1996) Income, residential location, and mode choice. *J Urban Econ* 40(1):84–99
- Dökmeci V, Berköz L (2000) Residential-location preferences according to demographic characteristics in Istanbul. *Landsc Urban Plan* 48(1-2):45–55
- Eaton J, Kortum S, Kramarz F (2004) Dissecting trade: firms, industries, and export destinations. *Am Econ Rev* 94(2):150–154
- Engelbreten Ø, Næss P, Strand A (2018) Residential location, workplace location and car driving in four Norwegian cities. *Eur Plan Stud* 26(10):2036–2057
- 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
- Fuchs VR (1986) His and hers: gender differences in work and income, 1959–1979. *J Labor Econ* 4(3):S245–S272
- Fujii S, Gärling T (2003) Application of attitude theory for improved predictive accuracy of stated preference methods in travel demand analysis. *Transport Res Part A: Policy Pract* 37(4):389–402
- García-López M-À (2012) Urban spatial structure, suburbanization and transportation in Barcelona. *J Urban Econ* 72(2-3):176–190
- Gerwe O (2021) The Covid-19 pandemic and the accommodation sharing sector: effects and prospects for recovery. *Technol Forecast Soc Change* 167:120733
- Green R, Hendershott PH (1996) Age, housing demand, and real house prices. *Regional Sci Urban Econ* 26(5):465–480
- Guidon S, Wicki M, Bernauer T, Axhausen K (2019) The social aspect of residential location choice: on the trade-off between proximity to social contacts and commuting. *J Transp Geogr* 74:333–340
- Heblisch S, Redding SJ, Sturm DM (2020) The making of the modern metropolis: evidence from London. *Q J Econ* 135(4):2059–2133
- Hong Y, Lyu X, Chen Y, Li W (2020) Industrial agglomeration externalities, local governments' competition and environmental pollution: evidence from Chinese prefecture-level cities. *J Clean Prod* 277:123455
- Huebner BM, Pleggenkuhle B (2015) Residential location, household composition, and recidivism: an analysis by gender. *Justice Q* 32(5):818–844
- Huu Phe H, Wakely P (2000) Status, quality and the other trade-off: towards a new theory of urban residential location. *Urban Stud* 37(1):7–35
- Jang S, Yi C (2021) Imbalance between local commuting accessibility and residential locations of households by income class in the Seoul Metropolitan Area. *Cities* 109:103011
- Kluger AN (1998) Commute variability and strain. *J Organ Behav: Int J Ind, Occup Organ Psychol Behav* 19(2):147–165
- Lai Y, Lv Z, Chen C, Liu Q (2022) Exploring employment spatial structure based on mobile phone signaling data: the case of Shenzhen, China. *Land* 11(7):983
- Lee BH, Waddell P, Wang L, Pendyala RM (2010) Reexamining the influence of work and nonwork accessibility on residential location choices with a microanalytic framework. *Environ Plan A* 42(4):913–930
- Levinson D (2008) Density and dispersion: the co-development of land use and rail in London. *J Econ Geogr* 8(1):55–77
- Li H, Campbell H, Fernandez S (2013) Residential segregation, spatial mismatch and economic growth across US metropolitan areas. *Urban Stud* 50(13):2642–2660
- Liu Y, Tang Y (2021) Epidemic shocks and housing price responses: evidence from China's urban residential communities. *Regional Sci Urban Econ* 89:103695
- Loa P, Hossain S, Mashrur SM, Liu Y, Wang K, Ong F, Habib KN (2021) Exploring the impacts of the COVID-19 pandemic on modality profiles for non-mandatory trips in the Greater Toronto Area. *Transp policy* 110:71–85
- Lyons G, Chatterjee K (2008) A human perspective on the daily commute: costs, benefits and trade-offs. *Transp Rev* 28(2):181–198
- Melia S, Chatterjee K, Stokes G (2018) Is the urbanisation of young adults reducing their driving? *Transport Res part A: Policy Pract* 118:444–456
- Næss P (2006a) Accessibility, activity participation and location of activities: exploring the links between residential location and travel behaviour. *Urban Stud* 43(3):627–652
- Næss P (2006b) Urban structure matters: residential location, car dependence and travel behaviour. Routledge
- Næss P, Strand A, Wolday F, Stefansdottir H (2019) Residential location, commuting and non-work travel in two urban areas of different size and with different center structures. *Prog Plan* 128:1–36
- Ni L, Wang XC, Chen XM (2018) A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data. *Transport Res part C: Emerg Technol* 86:510–526
- Pagliara F, Preston J, Simmonds D (2010) Residential location choice: models and applications. Springer Science & Business Media

- Peng C, Song M, Han F (2017) Urban economic structure, technological externalities, and intensive land use in China. *J Clean Prod* 152:47–62
- Phithakkittukoon S, Smoreda Z, Olivier P (2012) Socio-geography of human mobility: a study using longitudinal mobile phone data. *PLoS one* 7(6):e39253
- Portnov BA, Axhausen KW, Tschopp M, Schwartz M (2011) Diminishing effects of location? Some evidence from Swiss municipalities, 1950–2000. *J Transp Geogr* 19(6):1368–1378
- Rivas R, Patil D, Hristidis V, Barr JR, Srinivasan N (2019) The impact of colleges and hospitals to local real estate markets. *J Big Data* 6(1):1–24
- Sander W (2006) Educational attainment and residential location. *Educ Urban Soc* 38(3):307–326
- Schirmer PM, Van Eggermond MA, Axhausen KW (2014) The role of location in residential location choice models: a review of literature. *J Transp Land Use* 7(2):3–21
- Shin J, Tilahun N (2022) The role of residential choice on the travel behavior of young adults. *Transport Res part A: policy Pract* 158:62–74
- Stopher PR, Hartgen DT, Li Y (1996) SMART: simulation model for activities, resources and travel. *Transport* 23(3):293–312
- Taniguchi A, Fujii S, Azami T, Ishida H (2014) Persuasive communication aimed at public transportation-oriented residential choice and the promotion of public transport. *Transportation* 41(1):75–89
- Thorhauge M, Cherchi E, Rich J (2016) How flexible is flexible? Accounting for the effect of rescheduling possibilities in choice of departure time for work trips. *Transport Res Part A: Policy Pract* 86:177–193
- Tonne C, Adair L, Adlakha D, Anguelovski I, Belesova K, Berger M, Brelsford C, Dadvand P, Dimitrova A, Giles-Corti B (2021) Defining pathways to healthy sustainable urban development. *Environ Int* 146:106236
- Venter C, Vokolkova V, Michalek J (2007) Gender, residential location, and household travel: empirical findings from low-income urban settlements in Durban, South Africa. *Transp Rev* 27(6):653–677
- Verplanken B, Walker I, Davis A, Jurasek M (2008) Context change and travel mode choice: combining the habit discontinuity and self-activation hypotheses. *J Environ Psychol* 28(2):121–127
- Verplanken B, Wood W (2006) Interventions to break and create consumer habits. *J Public Policy Mark* 25(1):90–103
- Vilar-Compte M, Burrola-Méndez S, Lozano-Marrufo A, Ferré-Eguiluz I, Flores D, Gaitán-Rossi P, Teruel G, Pérez-Escamilla R (2021) Urban poverty and nutrition challenges associated with accessibility to a healthy diet: a global systematic literature review. *Int J Equity Health* 20:1–19
- Waddell P (2002) UrbanSim: modeling urban development for land use, transportation, and environmental planning. *J Am Plan Assoc* 68(3):297–314
- Wan L, Tang J, Wang L, Schooling J (2021) Understanding non-commuting travel demand of car commuters—Insights from ANPR trip chain data in Cambridge. *Transp Policy* 106:76–87
- Wang Y, de Almeida Correia GH, van Arem B, Timmermans HH (2018) Understanding travellers' preferences for different types of trip destination based on mobile internet usage data. *Transport Res Part C: Emerg Technol* 90:247–259
- Xu F, Li Y, Jin D, Lu J, Song C (2021) Emergence of urban growth patterns from human mobility behavior. *Nat Comput Sci* 1(12):791–800
- Yan L, Wang D, Zhang S, Xie D (2019) Evaluating the multi-scale patterns of jobs-residence balance and commuting time–cost using cellular signaling data: A case study in Shanghai. *Transportation* 46:777–792
- Yang H, Fu M, Wang L, Tang F (2021) Mixed land use evaluation and its impact on housing prices in Beijing based on multi-source big data. *Land* 10(10):1103
- Ye R, De Vos J, Ma L (2020) Analysing the association of dissonance between actual and ideal commute time and commute satisfaction. *Transport Res part A: policy Pract* 132:47–60
- Yu L, Zhao P, Tang J, Pang L, Gong Z (2023) Social inequality of urban park use during the COVID-19 pandemic. *Humanities Soc Sci Commun* 10(1):1–11
- Zhang J, Hayashi Y, Frank LD (2021) COVID-19 and transport: findings from a world-wide expert survey. *Transp Policy* 103:68–85

- Zhao P, Gao Y (2023) Discovering the long-term effects of COVID-19 on jobs–housing relocation. *Humanities Soc Sci Commun* 10(1):1–17
- Zhuge C, Shao C, Gao J, Dong C, Zhang H (2016) Agent-based joint model of residential location choice and real estate price for land use and transport model. *Comput Environ Urban Syst* 57:93–105

Acknowledgements

This study was supported by the National Natural Science Foundation of China (41925003, 42130402), and Shenzhen science and technology program (JCYJ20220818100810024, KQTD20221101093604016).

Author contributions

YC: writing—conceptualisation, original draft, methodology and formal analysis; PZ: supervision, writing—review and editing and funding acquisition; LL: writing—review and editing and formal analysis; JL: formal analysis; MG: methodology; YD: investigation; ZS: data curation; SY: visualisation; XD: visualisation.

Competing interests

The author declare no competing interests.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors. To protect personal information privacy, the mobile phone numbers of subscribers were anonymized by the mobile phone operator inside their premises and anonymized mobile signalling data were never transferred outside of the operator's system. Moreover, mobility data used in this study were aggregated according to time, space, and user attributes. This means that the analysis never singled out identifiable individuals.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-024-02678-2>.

Correspondence and requests for materials should be addressed to Pengjun Zhao.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024