Transport Policy 173 (2025) 103789

Contents lists available at ScienceDirect

Transport Policy

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

Modelling residential relocation behaviour combining passive revealed preference data and stated preference survey data

Yu Wang [a], Thomas O. Hancock [a], Yacan Wang [b], Charisma Choudhury [a,*]

[a] Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK
[b] School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, China

ARTICLE INFO

Keywords:
Residential location
SP-RP data combining
Long-term impact of pandemic

ABSTRACT

Understanding how various factors shape residential relocation is crucial for effective infrastructure planning and policy. Yet, existing revealed preference (RP) datasets often lack essential demographic or dwelling details, while stated preference (SP) surveys are prone to hypothetical bias and behavioural incongruence. To fill in this gap, this study presents a residential relocation choice model that combines residential location data derived from passively generated public transport smart cards of 82,720,872 users and SP data from 971 respondents (8739 observations) in Beijing, China. Both types of data were generated or collected in the backdrop of the COVID-19 pandemic, which led to higher-than-usual residential relocations in Beijing. The integrated approach, which accounts for the scale difference between the two datasets, reveals a strong preference for city-centre locations. But higher infection risks increase the likelihood of moving away from crowded areas, whereas flexible work-from-home policies lower the inclination to relocate to the centre. These findings quantify how different pandemic-related factors alter traditional relocation drivers. The results can guide policymakers in designing more resilient housing and transport policies, especially under future disruptions like pandemics. Moreover, the data-fusion framework offers a replicable strategy for researchers and planners seeking to capture both real-world behaviours and hypothetical scenarios in residential location studies.

1. Introduction

Understanding the relative impact of different factors on residential location choice and relocation behaviour is crucial for developing corresponding urban planning strategies (Tao et al., 2023; Krizek, 2003). Residential relocation choice models, which quantify the sensitivity to different spatial and transport network attributes and the underlying heterogeneity among the preferences of different individuals/households, are useful for planners to predict the changes in demand and mobility requirements in a given location. This can help planners and policy makers in designing the infrastructure and adopting appropriate demand management policies (Robbennolt et al., 2023; de Palma et al., 2022; Conway et al., 2020). The need to develop residential relocation models in greater detail has become even more critical in the aftermath of COVID-19, where restrictive policy measures to control crowding levels on public transportation, increased flexibility to work from home, the rise of active travel modes, etc., have significantly influenced traditional stable travel behaviour. Consequently, there has been a substantial shift in residential location and mobility preferences in many parts of the world (e.g. Shakib et al., 2022; Ghosh et al., 2020; Chen et al., 2023).

Despite its importance, modelling residential relocation decisions remains a relatively under-researched topic given that it is typically an infrequent choice (Haque et al., 2019; Eluru et al., 2009). Previous studies on residential relocation have mostly used retrospective RP surveys (see Table 1), which can be prone to omission bias, rounding bias and other measurement errors. RP panel surveys, which record residential location records over the years, are regarded as a more dependable source of relocation data, but it has been acknowledged that it is often difficult to get a long panel of real-world observations of residential relocation (Shakib et al., 2022). Further, the residential relocation data on many of the long panels (e.g. British Household Panel Survey) are prone to biases as those who change residential locations are the ones more likely to drop out from such panels (Haque et al., 2019).

This article is part of a special issue entitled: Travel Behavior and Demand published in Transport Policy.

* Corresponding author.
E-mail addresses: Y.Wang4@leeds.ac.uk (Y. Wang), t.o.hancock@leeds.ac.uk (T.O. Hancock), ycwang@bjtu.edu.cn (Y. Wang), c.f.choudhury@leeds.ac.uk (C. Choudhury).

https://doi.org/10.1016/j.tranpol.2025.103789
Received 1 June 2024; Received in revised form 22 August 2025; Accepted 27 August 2025
Available online 28 August 2025

Although recent studies have started to utilise passively generated revealed RP data (passive-RP), such as smart card and mobile phone data, these datasets often lack detailed demographic information and specific dwelling attributes (Chen et al., 2023; Wiig et al., 2020; Huang et al., 2018). Furthermore, the granularity of these data can be coarse, making it difficult to extract detailed location attributes (Chen et al., 2023; Huang et al., 2019). Stated preferences (SP) surveys have been used as an alternate approach to collect data on the relocation choices of individuals in different hypothetical scenarios, constituting different levels of potential influencing factors alongside the characteristics of the decision maker. However, these are prone to hypothetical bias and behavioural incongruence (Hensher, 2010). For example, Earnhart (2002) used a combined SP-RP model to address this issue. However, the survey asked people to retrospectively report RP location choice and residential attributes and can be prone to biases due to the recalled nature of the data.

To overcome the limitations of relying on a single source of data, researchers have proposed joint SP-RP modelling, an approach that integrates SP and RP data within one framework (Ben-Akiva and Morikawa, 1990; Hensher et al., 1998). This method harnesses the complementary strengths of each data source by 'calibrating' the hypothetical scenarios presented in SP with actual, real-world decisions from RP (Hensher, 2012). Consequently, the SP part supplies support to infer behaviours that single RP data cannot capture (for example, people's responses to new policy scenarios or transport modes) (Guzman et al., 2021; Helveston et al., 2018). Moreover, combining RP and SP data has repeatedly been shown to enhance model robustness and reliability, yielding greater statistical efficiency, improved parameter

Table 1
Previous studies on residential relocation.

References	Covid-19 context	District	SP data	RP data	Variables considered
Tayyaran & Khan (2007)	×	Ottawa, Canada	Stated choice experiment	Retrospective self-report	Residential, employment, and socioeconomic characteristics.
Earnhart (2002)	×	Long Island Sound, USA	Stated choice experiment	Retrospective self-report	Natural feature; House attributes; Flooding risk
Tran et al. (2017)	×	Hanoi, Vietnam	Household Interview Survey-expectations about housing type and location in future (not SP experiment)	Retrospective self-report	Land-use variables; Household attributes; Social demographic characteristics; Future expectation (not a joint SP-RP model)
Chen et al. (2023)	✓	Pearl River Delta, China	×	Passive data-Mobile phone GPS data	Average social demographic characteristics of districts; Spatial characteristics of the district
Zhao and Gao (2023)	✓	Beijing, China	×	Passive data-Mobile phone signalling data	Income; Commuting characteristics; COVID-19 infection
Huang et al. (2018)	×	Beijing, China	×	Passive data-Smart card data	Commuting time
Thomas et al. (2016)	×	Great Britain	×	Retrospective self-report -Cross-sectional consumer survey	Social demographic characteristics; Neighbourhood attributes; Residential duration
Ettema & Nieuwenhuis (2017)	×	Hague, Netherland	×	Retrospective self-report	Attitudes towards different travel modes; Travel mode accessibility; Social demographic characteristics; Built environment attributes
Wolday & Böcker (2023)	✓	Oslo and its surrounding municipalities, Norway	×	Retrospective self-report	Social demographic characteristics; Transportation conditions; House attributes; Built environment variables
Author(s)	Covid-19 context	District	SP data	RP data	Variables considered
Bhat & Guo et al. (2007)	×	Alameda County, USA	×	Retrospective self-report -San Francisco Bay Area Travel Survey (BATS)	Zonal size and density; Zonal land-use structure variables; Regional accessibility; Commute-related variables; Local transportation network; Zonal demographics and housing cost; Zonal ethnic composition
Stawarz et al. (2022)	✓	Germany	×	Time-series data of annual inter-county migration flows	Age; District
Kim (2006)	×	–	Stated choice experiment	×	House price; Travel time/cost to work/shop; Noise; Access to park/school; Dwelling size; Income
Krueger et al. (2019)	×	Sydney, Australia	Stated choice experiment	×	Mobility tools (focus on self-driving car); Rent; House attributes; Neighbourhood attributes
Liao et al. (2015)	×	Utah, USA	Stated choice experiment	×	Distance to work; Distance to different facilities; Neighbourhood attributes; Rent; Social demographic characteristics; Attitudes towards Nature and social environment/travel mode/ Policies
Tillema et al. (2010)	×	Netherland	Stated choice experiment	×	Toll; Number of bedrooms; Location and environment; Commuting time/cost
Stokenberga (2019)	×	Bogota', Colombia	Stated choice experiment	×	Family network; Type of housing; Tenure and Price; Time to CBD; Size
Balbontin et al. (2015)	×	Chile	Stated choice experiment	×	Rent; Public transportation condition; Distance to parks/services/cultural area; House attributes; Cleanliness of the street

significance, and a better overall fit than models relying on one data source alone (Guzman et al., 2021; Whitehead and Lew, 2020). For instance, Guzman et al. (2021) label a combined SP-RP mode choice model as a more robust tool for travel behaviour analysis and forecasting. Whitehead and Lew (2020) document notable gains in econometric efficiency, producing more precise estimates and tighter confidence intervals. These properties are critical for policy simulations, as they minimise the biases or peculiarities inherent in any single dataset. Furthermore, joint SP-RP frameworks facilitate validation and calibration of behavioural assumptions, since the real-world RP data offer an external check on SP-based predictions, providing external validity. Buckell and Hess (2019), for example, demonstrate that adding even a modest amount of real-market RP data to an SP-focused policy model substantially improves the accuracy of forecasts, aligning key metrics (e.g., elasticities, willingness-to-pay) more closely with actual observed behaviour. This, in turn, gives policymakers increased confidence in the realism and applicability of the resulting forecasts.

Based on these insights, this study combines passively generated location data from public transport smart cards (referred to as passive-RP in the rest of the paper) with SP survey data to fill in the research gaps and get deeper insights into residential relocation and commuting patterns in Beijing, China. Residential location data derived from public transport smart cards from 82,720,872 users and SP data from 971 respondents (8739 observations) are used in this regard. Both datasets are generated/collected in the backdrop of the Covid-19 pandemic, which led to higher-than-usual residential relocations in Beijing. A SP-RP combined model is constructed and accounts for the potential scale difference between the coefficients of the models estimated using only the RP and SP datasets.

The rest of this paper is structured as follows: Section 2 proposes the hypotheses. Section 3 outlines the study design and methodology. Section 4 presents the results. Section 5 discusses the results, proposes policy implications and highlights limitations. Finally, section 6 summarises the findings and suggests directions for future research.

2. Hypotheses

Previous research on residential relocation mainly concentrated on identifying the factors related to the attractiveness of the new location that influence households' decisions when relocating (Liao et al., 2015; Balbontin et al., 2015). These works typically focus on the improvements in attributes of the dwelling, the built environment and travel-related factors associated with the new neighbourhood (Kleinhans and Kearns, 2013; Van Acker and Witlox, 2009). For example, Borgers and Timmermans (1993) proposed a framework to analyse residential relocation from 3 aspects: dwelling attributes (reasonable cost, better environment, etc), the transportation system (higher frequency of public transport services, availability of public transport facilities, etc.) and commuting attributes (reduction in commuting time and commuting cost, etc.). Some studies have expanded the scope to include the socioeconomic and other characteristics of the decision makers - their attitudes toward relocation and job-related considerations (e.g. Ettema and Nieuwenhuis, 2017; Kim, 2006). However, few frameworks pay attention to the interaction role of exogenous social factors on relocation behaviour, such as travel mode availability (Krueger et al., 2019), policies (Tillema et al., 2010), etc. It may be noted that the need to include the sensitivity towards different exogenous policy factors has become crucial in the context of COVID-19 where restrictive policies like maximum level of crowding on public transport and flexible working policies (like the requirement or option to work from home) have potentially played a substantial role in the residential relocation decisions (Hensher et al., 2021). Therefore, this study examines not only the effects of housing attributes, built environment and neighbourhood attributes on relocation choices, but also the interactions of pandemic-related factors and policy-related factors with these traditional determinants. For clarity, attributes can change either way in relocation decisions (e.g. renting price up or down). Below, we use the positive changes of attributes and propose hypotheses accordingly, such as longer commuting time, better access to living facilities, presence of restrictive PT policy (compared with 'without any restrictive policy'), etc.

We first discuss the factors that may negatively affect relocation choice. The increase in housing/renting prices, for instance, can create financial strain, prompting individuals to seek more affordable living arrangements (Balbontin et al., 2015; Kim, 2006). In addition to price considerations, increasing commuting time and fare are critical cost-related factors that can push commuters to relocate (Huang et al., 2018). Moreover, these variables will have new impacts during the pandemic. Economic pressures are especially acute during a pandemic when individual income stability may be influenced (Zhao and Gao, 2023; Qian and Fan, 2020). Longer commuting times not only reduce personal time but also increase exposure to health risks associated with public transportation during a pandemic (Harrington and Hadjiconstantinou, 2022; Ando et al., 2021). The increased transportation costs add to the financial burden, making relocation to areas with better transportation options a more attractive choice. Furthermore, residential decisions are also influenced by pandemic-related risks such as the infection rate in a given district and the duration for which the district may be perceived or labelled as having a medium/high risk. The perceived risk of staying in a heavily affected area can be a strong motivator for relocation to areas with fewer impacts of the pandemic (Yıldırım et al., 2021). Meanwhile, population density is a potential factor: higher population density areas make social distancing more difficult and may increase infection risk, making them less attractive during a pandemic.

Therefore, this study proposes the following hypothesis.

H1. Cost/negative factors, including … [ ] …, will decrease the utility to choose that area as the residential location (and trigger relocation).

(a) … [increase in price] …
(b) … [increase in commuting time/fare] …
(c) … [longer duration of high risk during the pandemic] …
(d) … [higher population density] …

Next, we discuss which factors may positively influence relocation choice. Various factors can make certain residential areas more attractive, thereby influencing relocation decisions. Firstly, the better availability and quality of public transportation are crucial factors. Efficient and reliable public transport systems reduce commuting time and costs, making daily travel more convenient (Ettema and Nieuwenhuis, 2017). Then, living facilities, such as access to shops, also enhance the attractiveness of a location. Areas with comprehensive living facilities provide a higher quality of life and greater convenience, making them desirable destinations for those looking to relocate (Krueger et al., 2019). Moreover, open spaces and green areas contribute to the appeal of a residential location (Tillema et al., 2010; Earnhart, 2002). These factors are not only important under normal contexts but also play an important role in the context of the pandemic. Better transportation conditions are especially important during a pandemic, where minimising time spent in transit can reduce exposure to health risks (Wolday and Böcker, 2023). Access to parks and green spaces can improve mental and physical well-being, which is particularly important during times of heightened stress during a pandemic (Venter et al., 2020). In addition, city centres often offer better access (e.g. within walking or cycling distance from home) to employment opportunities and essential services, making them attractive options for relocation during a pandemic (Chen et al., 2023). The convenience and accessibility of city centres can significantly enhance the quality of life and mitigate some of the challenges posed by the pandemic.

Thus, this study proposes the following hypothesis.

H2. A better condition of residence, including … [ ] …, increase the utility to choose that area as the residential location (and trigger relocation) … [proximity to city centre] …

(b) … [better access to public transport systems] …
(c) … [better access to living facilities] …
(d) … [better access to open space and green spaces] …

Moreover, pandemic-related factors and policy-related factors may have an interaction effect on the above relationships. Firstly, pandemic-related factors can moderate the relationship between living facilities and residential choice behaviour. While more living facilities generally enhance the attractiveness of a location, the higher risk of infection may weaken this relationship. More medical, educational, and shopping facilities attract more people, causing an increase in the likelihood of infection. Their appeal may reduce as people prioritise safety over convenience due to the increased risk of infection (Dryhurst et al., 2022). Conversely, the availability of open space and green areas can become even more critical. The risk of infection can intensify the positive impact of green spaces on residential choice behaviour (Berdejo-Espinola et al., 2021).

Additionally, policies can influence preferences, such as public transportation restrictions during a pandemic, which can intensify the relationship between commuting time and residential choice behaviour. When public transportation options are limited or considered unsafe, longer commuting times become even more burdensome, influencing individuals to relocate closer to their workplaces or to areas with more reliable transportation options (Christidis et al., 2022). Finally, the option to work from home can weaken the effects of how close a location is to the workplace and the city centre. This reduced dependence on physical proximity to the workplace allows for greater flexibility in choosing residential locations based on other factors such as living conditions, personal preferences, etc.

Therefore, we propose the following hypotheses.

H3. Rising of risk of infection will … [ ] … the relationship between … … and residential choice behaviour during a pandemic.

… [weaken]… … (better access to living facilities) …
(b) … [intensify]… … (better access to open space and green spaces) …

H4. PT restriction will intensify the relationship between longer commuting time and residential choice behaviour during a pandemic.

H5. Working from home will weaken the relationship between living in the city centre and residential choice behaviour during a pandemic.

3. Study design and data

In this study, we use both passive-RP data and SP survey data. The SP data, while having a smaller sample, enables us to present different hypothetical scenarios with varying levels of attributes and different policy scenarios. It also allows us to collect individual-level socio-demographic variables. On the other hand, the passive-RP data have a large sample size and reflect the actual choices of commuters. Though it includes aggregate level spatial and transport network attributes (e.g. built environment attributes), it lacks individual-level socio-demographic information. By combining the strengths of both data sources, an SP-RP joint model was constructed to provide more comprehensive insights.

3.1. Passive-RP data

The automated fare collection (AFC) system was introduced to Beijing's public transport in 2006. Passengers are required to use their smart cards at both entry and exit points, generating a pair of time-stamped trip records per journey. Utilising the unique ID on each card allows for the longitudinal tracking of subway usage patterns (Wang et al., 2020). Table 2 displays an example of the data extracted from these smart card data sources for this study.

The data source includes smart card data from 82,720,872 users (including one-way ticket users) between January 2020 and October 2020. This period covered the first and second waves of the pandemic in Beijing. Using the method from Huang et al. (2018), the home stations and work stations of subway commuters in January (before the pandemic) and in October (after the pandemic) are identified.

Firstly, commuters are identified based on a minimum travel frequency of three days per week pre-pandemic. Commuting trips are then identified by a 6-h gap between boarding and the previous trip's alighting time. It is assumed that only trips completed on the same day accurately capture return commutes, where inter-day travels are not identified as commuting trips. Therefore, trips where the return trips do not occur on the same day as the previous trips are excluded. For each commuter, for each time period, the most frequently visited origin station(s) during these return commutes are designated as the work station (s), while the corresponding most frequently visited destination station (s) are designated as the home station(s). This process ensures that only regular travel patterns are analysed, focusing on the predominant stations that define each commuter's work and residential locations. For detailed information about the method, the readers are referred to Huang et al. (2018).

This study focuses on commuters whose home and work locations are identifiable in both January and October 2020. Individuals whose locations are identifiable in January but not in October are excluded, as their absence may stem from various factors—such as remote work, loss of smart card, or relocation from Beijing—which cannot be definitively determined using smart card data alone. Based on this, we can obtain two pairs of home stations and work stations for each commuter before and after the pandemic. Consequently, this allows us to capture the trends in residents' relocation, commuting patterns and the attributes of home stations in each period.

Following the above data analysis process results in the identification of 425,439 commuters travelling to/from 340 distinct subway stations (out of 475 total stations). We further excluded commuters with multiple home or work stations, as they may reside or work in different locations while still meeting the criteria for commuting origin–destination patterns. For instance, a commuter travelling four days a week might split their time between their own home and their parents' home. For such cases, it is not possible to assign a unique set of commuting and residential characteristics to that single ID. After this additional filtering, we obtain 339,024 commuters with single home-work station pairs in both periods. Results revealed that 33.9 % of the commuters in our sample changed their home stations after the pandemic compared with before the pandemic (a proxy for their residential location). According to the 5th Beijing Comprehensive Transportation Survey (2016), Beijing's subway users are mainly young travellers, with most residing in rental housing, thereby exhibiting higher mobility. For instance, Huang et al. (2018) reported that over a span of seven years, 83.62 % of the subway commuter sample had relocated either their residence, workplace, or both. While we lack official Beijing statistics on pandemic-related moves, our survey respondents retrospectively reported high mobility during the pandemic as well. The data are presented in the next section.

Table 2
An example of smart card data.

Card ID	Check-in time	Origin station	Check-out time	Destination station
1000751085xxxxxx	20200130007200	9430	104207	0104
1000751085xxxxxx	20200201082500	9430	100526	0104
1000751017xxxxxx	20200130083600	9430	91301	0105

3.2. SP data

Each participant completed the questionnaire, which consisted of three parts: an SP experiment with 9 choice scenarios, a reported behaviour of whether or not the individual has relocated since the beginning of the COVID-19 pandemic (defined here as January 2020) and a section for reporting socio-demographics, commuting-related variables and the attributes of their current home.

In the SP, the respondents were asked to imagine a new pandemic breakout, and during the breakout, the participants were asked to choose a home between two alternatives. The attributes included the following, and the full list and variable descriptions are in Table 4.

(a) Traditional residential relocation variables:
- Dwelling attributes: such as the proportion of rent to monthly salary, whether it involves sharing the accommodation with non-family members;
- Neighbourhood attributes: such as whether the location is in the city centre or not, distance to public transportation, distance to essential amenities, distance to open spaces, and population density;

(b) Pandemic-related variables: the risk of infection;
(c) Policy-related variables: whether the company allows remote work and whether the government would restrict public transportation usage.

An efficient experimental design was employed to generate 18 scenarios, divided into two blocks. Prior to finalising the questionnaire, a pretest was conducted with 50 pilot respondents. The resulting estimation outcomes informed adjustments to the choice scenario design, question phrasing, and variable selection. An example choice scenario is shown in Fig. 1.

In addition to the choice experiment, participants were required to recall whether they had changed their residence between January and October 2020 (the period overlapping the passive-RP data). They were also asked to report attribute values of their current housing conditions corresponding to the variables used in the choice experiment, such as the distance to public transportation. Their reports served as a reference base in model construction. Finally, they were asked to provide

Table 3
Description of sampled commuters.

Socio-demographic attributes		Sample (%)	Beijing Census[a]
Gender	Male	47.17	50.9 %
	Female	52.83	49.1 %
Age	18–24	25.85	The main commuters on the Beijing subway are young people aged 20 to 40 [b].
	25–34	44.28	
	35–44	24.30	
	45–54	4.74	
	55 and above	0.82	
Education level	Below undergraduate degree	18.13	–
	Undergraduate degree	46.65	
	Above undergraduate degree	35.22	
Income	Less than ¥100k[c]	30.07	In 2023, the annual per capita disposable income of residents in the entire city was ¥81,752.
	¥100k-¥200k	38.83	
	¥200k-¥300k	22.66	
	¥300k-400k	6.49	
	Above ¥400k	1.96	

[a] Beijing's 2023 Statistical Bulletin on National Economic and Social Development. https://tjj.beijing.gov.cn/tjsj_31433/sjjd_31444/202403/t20240319_3594001.html.
[b] Report of the 5th Beijing Comprehensive Transportation Survey, 2016.
[c] 1 CNY≈0.14 USD.

Table 4
The variables used for the model.

Categories	Variables	Description
SP model component		
Pandemic related variables	$X_{nt}^{SP_{RI}}$: Risk of infection	The number out of every 100,000 people who would get infected when coming in contact with an infected person. (proportion)
Policy related variables	$X_{nt}^{SP_{WFH}}$: Company-Work from home	How many days employees can work from home during a week.
	$X_{nt}^{SP_{PTR}}$: Government-PT restriction	Whether there are public restriction measures the government takes to control the pandemic. PT will not provide service at stations near the place of residence. (1/0)
House attributes	$X_{int}^{SP_{R}}$: Rent per month	Amount of rent in CNY per month.
	$X_{int}^{SP_{S}}$: Share	Whether the house is shared with other renters. (1/0)
BE attributes and neighbourhood attributes	$X_{int}^{SP_{PT}}$: Distance to public transportation	Distance from home to public transit stations/stops.
	$X_{int}^{SP_{EA}}$: Distance to essential amenities	Distance from home to destinations such as shops, restaurants, public libraries and schools.
	$X_{int}^{SP_{os}}$: Distance to open space and green spaces	Distance from home to park and public open space.
	$X_{int}^{SP_{PD}}$: Population density	The number of residents per square kilometre.
Individual variables	$X_{int}^{SP_{SDV}}$: Social demographic variables	Gender, age, education level, income.
Choice	Y_{int}^{SP}: Choice of house	Choice between houses 1 and 2
Passive-RP model component		
Pandemic related variables	$X_{i'n'}^{RP_{RD}}$: Risk duration	Number of days the station was in a zone deemed high or medi-um risk[c].
House attributes	$X_{i'n'}^{RP_{HP}}$: House price	Average house price in CNY[b].
Built environment and neighbourhood attributes	$X_{i'n'}^{RP_{CF}}$: Commuting fare	Commuting fare[c]
	$X_{i'n'}^{RP_{LUM}}$: Land use mix	Degree of mixing of different land uses within an area, such as residential, commercial, and public spaces, etc.[b],[a]
	$X_{i'n'}^{RP_{BS}}$: Bus stops	Number of bus stops[b]
	$X_{i'n'}^{RP_{ML}}$: Metro lines	Number of metro stations[b]
	$X_{i'n'}^{RP_{M}}$: Mall	Whether the station is integrated as a transport hub with a mall[c].
	$X_{i'n'}^{RP_{MF}}$: Medical facility	The number of medical service facilities, such as hospitals, pharmacies, etc.[b]
	$X_{i'n'}^{RP_{EA}}$: Essential amenities	The number of living service facilities, such as shops, laundry, restaurants, post office, etc.[b]
	$X_{i'n'}^{RP_{EF}}$: Educational facility	The number of schools educational facilities, such as school, training centres, university, etc.[b]
	$X_{i'n'}^{RP_{LF}}$: Leisure facility	The number of leisure activities places such as museums, galleries, etc.[b]
	$X_{i'n'}^{RP_{OF}}$: Outdoor facility	The number of outdoor and activity places, such as parks, gyms, etc.[b]
Passive-RP model component		
Choice	$Y_{i'n'}^{RP}$: Choice of home stations	For movers, the original station, the new station (after moving), and five randomly chosen stations from all other station alternatives;

(continued on next page)

Table 4 (*continued*)

Categories	Variables	Description
		For stayers, the original station, and five randomly chosen stations from all other station alternatives.
Shared SP and RP component		
Neighbourhood and built environment variables	$X_{int}^{SP_{CT}}, X_{i'n'}^{RP_{CT}}$: Commuting time	Commuting time (mins) to work by public transit.
	$X_{int}^{SP_{L}}, X_{i'n'}^{RP_{L}}$: Location	The location of the house. City centre or Suburbs (1/0)

[a] http://www.thinkstreetsmart.org/land-use-mix.html.
[b] Aggregate values over a radius of 1 km around the station.
[c] Station-specific variable.

information on their socio-economic and commuting attributes.

The target population of the survey was public transport commuters from Beijing. A web-based survey was conducted in October and November 2023, recruiting 1156 Beijing residents via the online survey platform Credamo. After excluding the non-PT commuters, 971 valid responses were retained (valid rate: 84 %). Descriptive statistics of the 971 respondents in the final sample are presented in Table 3. Among the respondents, 47.17 % were male and 70.13 % were young commuters (under 35 years old). This is representative of typical subway commuters (5th Beijing Comprehensive Transportation Survey, 2016). However, it should be noted that the income level of the sample was slightly higher than the average level in Beijing.

In addition, analysis of the retrospective reports of residential locations at different points in the pandemic revealed that 53 % moved between January 2020 and October 2020, 59 % between October 2020 and December 2022 and 63 % after December 2022.

3.3. Model specification

Combining RP and SP survey data has a long tradition in discrete choice modelling (Ben-Akiva and Morikawa, 1990). Effective joint modelling assumes that both data sets capture the same essential attributes and that individuals' perceptions of these attributes remain consistent across different choice scenarios. This synergy allows for using the same parameters across SP and RP data, based on the hypothesis that individuals make consistent trade-offs (Börjesson, 2008). In the context of this study, the parameters of commuting time and residential location serve as common variables between the SP and RP components, acting as a bridge. Due to limitations in data collection, the RP data do not capture certain interaction-related variables, such as policy factors like remote work instructions, which were unobservable in our dataset. Consequently, their effects are primarily examined through the stated preference (SP) data, where hypothetical scenarios can be more easily constructed. The variables used for modelling are shown in Table 4.

In SP-RP joint estimation, scale differences in unobserved error variances between RP data and SP data are one of the key considerations (Ben-Akiva and Morikawa, 1990; Hensher, 1994, 2012). To allow for scale differences between the two data sources, we incorporate separate scale parameters μ_{RP} and $\mu_{SP,}$ where the μ_{RP} is kept fixed to 1 for normalisation (Beck et al., 2017; Axhausen et al., 2006). The specifications of the joint model are listed below.

For the SP component, the specification of the utility U_{int}^{SP} can be expressed in general terms as:

$$U_{int}^{SP} = \mu_{SP}\left(V_{int}^{SP} + \varepsilon_{int}^{SP}\right) \tag{1}$$

Where μ_{SP} is the scale parameter, i is the alternative, n is the individual, and t is the choice task. ε_{int}^{SP} is the error term. The V_{int}^{SP} represents the systematic utility expressed in equation (2):

$$V_{int}^{SP} = \sum_{k=1}^{K} \beta_k^{SP} X_{1intk}^{SP} + \sum_{h=1}^{H} \beta_h^{SP} \left(\frac{X_{2inth}^{SP}}{B_{n,h}^{SP}}\right)^{\lambda_h^{SP}} X_{2inth}^{SP} \tag{2}$$

X_{1intk}^{SP} and X_{2inth}^{SP} are the value of alternative attributes and β_k^{SP} and β_h^{SP} are respective parameters. For the attributes X_{2inth}^{SP} including distance to public transportation, distance to essential amenities, distance to open space and green spaces, population density and commuting time, the

Categories	Variables	Housing option 1	Housing option 2
Pandemic related variables	Risk of infection	5,000 (5%)	
Policy	Company measurement:Work from home	1-4 days a week	
	Government measurement: PT restriction	Yes	
House attributes	Rent per month	20%	50%
	Commuting time	40-60 mins	20 mins or less
	Location	Suburb	City centre
	Type	Share	Share
Built environment attributes and neighbourhood attributes	Transportation	Less than 500m	1000-1500m
	Living facilities	Less than 500m	10 miles or more
	Open space and green spaces	2–3 km	2–3 km
	Population density	5000 or less	10000-15000
Your preferred house			

Fig. 1. An example choice scenario.

nonlinear effects are considered by incorporating an equation $\left(\frac{X^{SP}_{2inth}}{B^{SP}_{n,h}}\right)^{\lambda^{SP}_h}$. $B_{n,h}$ is the reference value for the attribute. We use the value of the attributes of their current home reported by the respondents as the base for the built environment attributes and commuting time. λ^{SP}_h is an elasticity parameter, giving the elasticity of the sensitivity to X^{SP}_{2inth} with respect to the reference value, with negative values for λ^{SP}_h the (absolute) sensitivity decreases with increases (relative to the base) in X^{SP}_{2inth}, with the opposite applying in the case of positive values of λ^{SP}_h. X^{SP}_{1intk} are other explanatory and control variables.

In addition, to test the effect of the interaction factors, we add some interactions into the utility function, such as the interaction of the distance to essential amenities and risk of infection.

The probability P^{SP}_{int} of subject n choosing an alternative i in task t is then given by:

$$P^{SP}_{int} = \frac{exp\left(\mu_{SP} V^{SP}_{int}\right)}{\sum_{j\in C} exp\left(\mu_{SP} V^{SP}_{jnt}\right)} \tag{3}$$

The likelihood L^{SP} is calculated by:

$$L^{SP} = \prod_{int} \left(P^{SP}_{int}\right)^{y^{SP}_{int}} \tag{4}$$

$y^{SP}_{int} = 1$ if the respondent n chooses i in task t and 0 otherwise.

In residential location choice contexts, decision makers are faced with an extensive range of alternatives. In RP-based studies, it is impossible to observe the exact choice set considered by each individual. Moreover, estimating location choice models using the full set of spatial alternatives theoretically available to each household is neither computationally practical nor behaviourally realistic, as it is highly unlikely that households evaluate every possible option (Zolfaghari et al., 2012). To address these challenges, sampling of alternatives has been introduced as an effective method to reduce the size of the choice set and, consequently, the estimation burden—while still preserving the model's explanatory power and yielding behaviourally meaningful results in studies involving large choice sets (Tsoleridis et al., 2022; Guevara and Ben-Akiva, 2013; McFadden, 1972). This study adopted this method to construct the choice set for the RP model. For movers, we include the original station, the new station (after moving), and five randomly chosen stations from all other observed home station alternatives. For stayers, the set is similar but without the new station. We tested different numbers of random alternatives, and the results remained relatively stable.

For the RP component, similar model specifications are used. The specification of the utility $U^{RP}_{i'n'}$ is expressed as:

$$U^{RP}_{i'n'} = \mu_{RP}\left(V^{RP}_{i'n'} + \varepsilon^{RP}_{i'n'}\right) \tag{5}$$

Where μ_{RP} is the scale parameter; i' is the alternative ($i' = 1$ to 77 for movers and $i' = 6$ for stayers); n' denotes the individual, and $\varepsilon^{RP}_{i'n'}$ is the error term.

The $V^{RP}_{i'n'}$ represents the utility expressed in equation (6):

$$V^{RP}_{i'n'} = \rho^{RP}_{i'} + \sum_{k'=1}^{K'} \beta^{RP}_{k'} X^{RP}_{1i'n'k'} \tag{6}$$

$\rho^{RP}_{i'}$ is the alternative-specific constants. We only assigned an ASC to the original station, representing individuals' initial preference. $X^{RP}_{1i'n'k'}$ are the explanatory variables in the corresponding utility function are the attributes of the commuter's home station in January and October, as well as the corresponding commuting characteristics and $\beta^{RP}_{k'}$ are respective parameters.

The probability $P^{RP}_{i'n'}$ of subject n' choosing an alternative i' is then given by:

$$P^{RP}_{i'n'} = \frac{exp\left(\mu_{RP} V^{RP}_{i'n'}\right)}{\sum_{j'\in C'} exp\left(\mu_{RP} V^{RP}_{j'n'}\right)} \tag{7}$$

The likelihood L^{RP} is calculated by:

$$L^{RP} = \prod_{i'n'} \left(P^{RP}_{i'n'}\right)^{y^{RP}_{i'n'}} \tag{8}$$

Then, the final log-likelihood for the joint model is then given by:

$$LL = log\left(L^{SP} L^{RP}\right) \tag{9}$$

The model coefficients are estimated by maximising this function. The Apollo choice modelling package (Hess and Palma, 2019) is used in this regard.

4. Results

The results of the joint SP-RP model are shown in Table 5. The descriptions of the variables are listed in Table 4.

As observed in Table 5, the estimation results of the shared parameters indicate a significant preference among commuters for residential locations in city centre areas. This finding aligns with Chen et al. (2023) but contrasts with findings from Western contexts (where a trend towards suburban relocation has been reported during the pandemic (Stawarz et al., 2022; Zarrabi et al., 2021). A potential reason could be the fact that proximity to the city centre enables individuals to meet part of their mobility needs by means of active modes (walking and cycling), where the infection risk is less than public transport. Further, living in the city centre may reduce the inconvenience during public transport restrictions imposed during a pandemic. The coefficient for commuting time indicates a statistically significant negative elasticity parameter, which means that the longer the commuting time is compared to the reference time, the more the marginal utility decreases. This aligns with the shape of the value function in prospect theory, where sensitivity to changes decreases as the distance from the reference point increases. The shared parameters hence support H1b and H2a.

For the SP component, higher rent levels are found to serve as a significant factor in reducing the utility associated with a dwelling (supporting H1a). In addition, it is understandable that sharing accommodations with non-family members during the pandemic is a negative influencing factor. Commuters prefer renting alone to reduce the risk of infection. In contrast, there is a significant preference for residences nearer (i.e. shorter distance from) public transport stops, essential amenities, and open spaces, allowing for convenient travel and access to necessities in a pandemic context. Although the effect of population density is not significant, the sign of the coefficient is as expected, indicating a preference for living in less densely populated areas. Hence, the estimation results support H2b, c and d but not H1d.

However, the influence of interaction factors suggests that proximity to essential amenities and open spaces becomes less desirable as infection rates rise. A potential reason for this is that overcrowding in open spaces is a recurring problem in Beijing (Lee et al., 2014), and hence, due to the prevalence of high infection rates, they can be unsafe. Moreover, when government restrictions on public transportation are in place, longer commuting times significantly decrease utility for a location. This effect was particularly important given that our sample consisted of public transit commuters with limited alternative commuting options. Additionally, when longer durations of remote work is allowed, the preference for living close to the city centre is found to decrease. This is because typically, when someone can work from home, they can save the commute time altogether and potentially have more leisure time. Therefore, living close to the city centre becomes less appealing to them. In conclusion, H3a, H3c, H4 and H5 are supported, while H3b is not supported.

For the RP component, most of the results are consistent with the SP

Table 5
The results of the joint model.

Category	Variables	Estimate	Rob.t-ratio(0)
SP model component			
House attributes	$X_{int}^{SP_R}$: Rent per month	−0.2505	−10.65
	$X_{int}^{SP_S}$: Share	−0.2259	−4.44
BE attributes and neighbourhood attributes	$X_{int}^{SP_{PT}}$: Distance to public transportation	−0.3654	−3.45
	$X_{int}^{SP_{EA}}$: Distance to essential amenities	−0.4006	−2.76
	$X_{int}^{SP_{os}}$: Distance to open space and green spaces	−0.3833	−4.06
	$X_{int}^{SP_{PD}}$: Population density	−0.0042	−0.48
Interaction factors	$X_{int}^{SP_{EA}} \times X_{nt}^{SP_{RI}}$: Essential amenities × Risk of infection	0.0755	5.97
	$X_{int}^{SP_{os}} \times X_{nt}^{SP_{RI}}$: Open space and green spaces × Risk of infection	0.0400	7.05
	$X_{int}^{SP_{PD}} \times X_{nt}^{SP_{RI}}$: Commuting time_PT restriction	−0.0386	−4.62
	$X_{int}^{SP_{CT}} \times X_{nt}^{SP_{PTR}}$: Commuting time × PT restriction	−0.0124	−6.43
	$X_{int}^{SP_L} \times X_{nt}^{SP_{WFH}}$: Location × Work from home	−0.0655	−7.74
Elastic parameter	λ_h^{SP1}: Commuting time	0.3778	3.42
	λ_h^{SP2}: Distance to public transportation	−0.0147	−0.07
	λ_h^{SP3}: Distance to essential amenities	−0.0461	−0.20
	λ_h^{SP4}: Distance to open space and green spaces	−0.1785	−1.19
	λ_h^{SP5}: Population density	0.4598	1.93
Social demographic factors	$X_{int}^{SP_{SDV}}$: Male	0.1454	4.34
	$X_{int}^{SP_{SDV}}$: Education level_undergraduate degree	−0.0430	−0.99
	$X_{int}^{SP_{SDV}}$: Education level_Above undergraduate degree	0.0224	0.52
	$X_{int}^{SP_{SDV}}$: Age_18-24	−0.2269	−4.70
	$X_{int}^{SP_{SDV}}$: Age_25-34	−0.2124	−3.96
Social demographic factors	$X_{int}^{SP_{SDV}}$: Age_35-44	−0.4241	−6.07
	$X_{int}^{SP_{SDV}}$: Age_45-54	−0.2391	−2.64
	$X_{int}^{SP_{SDV}}$: Income_¥100k-¥200k	0.0142	0.38
	$X_{int}^{SP_{SDV}}$: Income_¥200k-¥300k	−0.1402	−2.84
	$X_{int}^{SP_{SDV}}$: Income_¥300k-400k	−0.1690	−2.33
	$X_{int}^{SP_{SDV}}$: Income_Above ¥400k	0.1118	1.57
Passive-RP model component			
ASC	$\rho_{i'}^{RP}$	2.6321	632.00
Pandemic related variables	$X_{i'n'}^{RP_{RD}}$: Risk duration	−0.0209	−8.68
Dwelling attributes	$X_{i'n'}^{RP_{HP}}$: House price	−0.1457	−48.83
Built environment and neighbourhood attributes	$X_{i'n'}^{RP_{CF}}$: Commuting fare	−0.2226	−60.61
	$X_{i'n'}^{RP_{LUM}}$: Land use mix	0.0185	0.81
	$X_{i'n'}^{RP_{BS}}$: Bus stops	−0.1026	−37.01
	$X_{i'n'}^{RP_{ML}}$: Metro lines	−0.1364	−20.83
	$X_{i'n'}^{RP_M}$: Mall	0.0769	9.29
	$X_{i'n'}^{RP_{MF}}$: Medical facility	−0.0412	−7.84
	$X_{i'n'}^{RP_{EA}}$: Essential amenities	0.0796	11.54
	$X_{i'n'}^{RP_{EF}}$: Educational facility	−0.0736	−16.22
	$X_{i'n'}^{RP_{LF}}$: Leisure facilities	0.0565	9.55

Table 5 (*continued*)

	$X_{i'n'}^{RP_{OF}}$: Outdoor facilities	−0.0936	−14.05
Shared SP and RP component			
Shared parameter	$X_{int}^{SP_{CT}}, X_{i'n'}^{RP_{CT}}$: Commuting time	−0.0069	−37.40
	$X_{int}^{SP_L}, X_{i'n'}^{RP_L}$: Location	0.4619	40.94
Scale parameter	RP	1.0000	–
	SP	1.9086	10.53
Goodness-of-fit		SP	RP
	LL(0)	−6057.41	−588,589.3
	LL (Final)	−4359.93	−318,254.41
		Whole model	
	LL(0)	−594,646.7	
	LL (Final)	−322,614.3	
	AIC	645,314.7	
	BIC	645,618.9	

findings. For example, high housing prices in the neighbourhood negatively influence the preference for a location. The length of periods marked as medium or high-risk for COVID-19 also reduces the attractiveness of those areas. In addition, the presence of more bus stations and more subway lines has the opposite effect. During the pandemic, travellers were cautious about areas with a high density of transit infrastructure, such as locations with numerous bus stops or multiple metro lines. This is likely because such areas tend to attract larger crowds and higher passenger turnover, increasing the perceived risk of virus transmission. In line with the findings of the SP model, essential amenities play a positive role in attracting relocation. However, the greater numbers of medical facilities and educational facilities reduce the preference for residential choice. The strict lockdown policies in Beijing, which could lead to entire blocks being locked down upon the discovery of a single case,[1] might explain why proximity to medical facilities could be seen as a disadvantage during the pandemic. For similar reason, when cases occur in a school or university where there are lots of vulnerable students, nearby residential areas would also be affected and locked down. Therefore, commuters choose to live away from these places to reduce the risk of being affected. Outdoor facilities have a significant negative effect. This aligns with the discussion in the SP section, where the influence of interaction factors suggests that proximity to essential amenities and open spaces becomes less desirable as infection rates rise. Although the specific variable forms are different, the results are consistent with the hypotheses tested and in general align with the findings from the SP component, indicating the robustness of the result.

In addition, as shown in the table, the SP submodel's LL improves from −6057.41 to −4359.93, while the RP submodel's LL improves from −588,589.3 to −318,254.41. Both improvements suggest that our submodels capture explanatory power beyond random choice. In conclusion, the combination of two datasets provides more insights than a single dataset alone and a more robust testing of the hypotheses (as summarised in Table 6).

5. Discussion and policy implication

The results of this study provide valuable insights into how urban planners and policymakers can better understand the new trends in residential relocation and commuting patterns in the context of a pandemic. Based on the findings, several policy implications are discussed.

[1] Beijing implemented a dynamic zero-COVID policy, which included strict regional lockdown measures. https://www.beijing.gov.cn/ywdt/gzdt/202205/t20220519_2716655.html.

Table 6
The support of the hypotheses.

Hypotheses	Variables	Influence		SP	RP
H1a	Increase in price	Negatively	Preference for residential choice	✓	✓
H1b	Increase in commuting time/ fare]			✓	✓
H1c	Longer duration of risk during the pandemic			–	✓
H1d	Higher population density			×	–
H2a	Proximity to city centre	Positively		✓	✓
H2b	Better access to public transport systems			✓	✓
H2c	Better access to essential amenities			✓	✓
H2d	Better access to open space and green spaces			✓	×
H3a	Increasing risk of infection	Weaken	Essential amenities	✓	–
H3b		Intensify	Open space and green spaces	×	–
H4	PT restriction	Intensify	Commuting time	✓	–
H5	Working from home	Weaken	Proximity to the city centre	✓	–

Note: √ supported; × not supported; – not tested.

During the pandemic, commuters often seek safer and more comfortable environments by relocating. However, housing costs remain a significant constraint, particularly for lower-income groups who may have fewer options and face higher pandemic-related risks. Policymakers should consider the affordability for low-income groups by providing more secure and affordable housing options, such as the affordable rental housing initiatives in Beijing. Additionally, improving public transportation infrastructure and systems to reduce commuting times and enhance efficiency is important, particularly under public transportation restrictions. Furthermore, the results show that the duration of an area's pandemic risk level is a significant factor that deters commuters from living in the area. Policymakers need to implement differentiated measures based on the specific pandemic situations of different areas to reduce the risk of virus transmission. Effective risk communication is essential to keep residents informed about the progress of the pandemic in different districts of the city and to reduce their perceived risks and fears. This helps prevent certain areas from being psychologically labelled as risky and unsafe by residents over the long term, which could negatively impact the area's local economy.[2]

The results from this study indicate a trend differing from that observed in many Western cases, with commuters showing a preference for central urban locations. Although we noted several advantages of the city centre during the pandemic, the downsides are equally evident. Since high-density city centres typically have more complex urban systems, they are more susceptible to the impacts of the pandemic and related preventive measures. The concentration of a larger population in these areas can also reduce the overall resilience of the system. Our results reflect these tensions: access to essential amenities is valued, yet close proximity to medical, education and large public transport hubs can reduce attractiveness during the pandemic. Therefore, during the pandemic, the accessibility of city-centre amenities and high population density may have mixed effects. Policymakers need to handle pandemic policies for different areas more cautiously and avoid a one-size-fits-all approach. Special care must be taken when designing measures for city centres. Policy should aim to keep the benefits of access while managing crowding and risk, for example, through local service provision, crowd control, and clear risk communication (Wang et al., 2024). In addition, commuters prefer living near public transportation stops and essential amenities. Therefore, urban planning should aim for the balanced development of various facilities across different areas to enhance accessibility and equity.

Regarding the findings related to interaction factors, as infection risk rises, residences near some essential amenities and open spaces become less attractive due to the increased flow of people and heightened infection rate. Policymakers should strengthen the management of these areas during pandemics to reduce infection risks and communicate this well to provide reassurance to the residents. This includes implementing time-based flow control, timely disinfection, and social distancing restrictions to maintain safety in high-flow areas. When there are restrictions on public transportation, the preference for housing with long commuting times significantly decreases. This underscores the importance of maintaining transportation supply during a pandemic. Policymakers should strive to keep public transportation operational and safe, ensuring that commuters can rely on it even during restrictive periods. Furthermore, increased flexibility of working remotely reduces the preference to choose a central location. Encouraging employers to adopt flexible work-from-home policies can reduce commuting and housing costs for employees, while also balancing population distribution across different districts. This can help mitigate the potential negative externalities of overpopulation in urban centres, which can create stress on the urban infrastructure, as well as posing the risk of increased transmission rates.

However, no study is without limitations. Although this paper analyses a range of influencing factors, this framework cannot incorporate all possible factors, such as attitudes, long-term planning and other socioeconomic variables. Future research should consider these factors to further explore the decision-making process of commuters. Additionally, due to the limitations of passive-RP data, modelling work is reliant on average characteristics near subway stations and cannot capture more detailed housing attribute variables or social demographic characteristics. Future studies will require a broader range of survey data to investigate more detailed information or incorporate more data sources, such as census data. In addition, inverse discrete choice modelling and latent demographics modelling are also useful ways to contribute to the enrichment of socio-demographic attributes for anonymous big datasets (Zhao et al., 2022; Bwambale et al., 2019).

6. Conclusion

This study empirically examines the change in residential location and commuting patterns of subway commuters in Beijing during the COVID-19 pandemic by combining passive-RP data with SP survey data. The analysis reveals that the pandemic has led to significant shifts in relocation trends, with a preference for city centre locations. However, this preference is weakened by the option to work from home. The study also finds that traditional influencing factors, such as the availability of essential amenities and outdoor spaces, have different impacts in the pandemic context. Specifically, while commuters generally prefer closer essential amenities and outdoor space, this preference diminishes as infection risks rise. The specific results are expected to help the city authorities better understand changes in residents' preferences and the subsequent new trend in housing and transport demand in the event of future pandemics.

It may be noted, though, that some findings are different from those from other countries and cities, highlighting that urban policymakers cannot simply adopt findings from other cases without considering their unique local contexts. This research underscores the need for

[2] For example, during the pandemic, Beijing's largest agricultural wholesale market Xinfadi and its surrounding areas were marked as high-risk district for two months, with all commercial activities suspended. Beijing residents avoided visiting that area for a prolonged period. https://baijiahao.baidu.com/s?id=1737611253634593744&wfr=spider&for=pc.

differentiated urban planning and policy responses that address the specific dynamics of each city during the breakout of a pandemic.

Further, the framework will be useful for planners and policy makers to consider beyond the pandemic. It highlights that in contexts where external factors can potentially result in a radical change in residential relocation and mobility behaviour (as was the case during the COVID-19 pandemic), it is important to understand the dynamic changes in preferences. Yet, in such contexts, it can be challenging to quickly conduct a large-scale survey and using multiple data sources can be the way forward. Combining different data sources can not only compensate for their shortcomings of each individual dataset but also offer more comprehensive insights. The analytical framework presented in this research for combining SP data with passively generated RP data to get the best of both worlds is expected to serve as a valuable reference for policymakers.

CRediT authorship contribution statement

Yu Wang: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Thomas O. Hancock:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Yacan Wang:** Writing – review & editing, Supervision, Funding acquisition, Data curation. **Charisma Choudhury:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

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.

Acknowledgement

Charisma Choudhury, Thomas Hancock and Yu Wang's time were supported by UKRI Future Leader Fellowship [MR/T020423/1] and [MR/Y034384/1]. Yacan Wang was supported by the Natural Science Foundation of China [grant number 72474018 and 72071017].

Appendix

The attributes and levels for the SP survey design are shown in Table A-1.

Table A-1
Attributes and levels for the SP survey design

Categories	Variables	Description	Level	Sources
Pandemic related variables	Risk of infection	The number out of every 100,000 people who would get infected when coming in contact with an infected person.	No pandemic, 5000 (5 %), 7500 (7.5 %), 10,000 (10 %)	Hess et al. (2022)
Policy related variables	Company measurement: Work from home	How many days the company allows working from home in a week	0, 1–4 days per week, 5 days per week	Delventhal et al. (2022); Currie et al. (2021)
	Government measurement: PT restriction	The measures the government takes to control the pandemic. PT is not provided at stations near the place of residence.	Yes No	
House attributes	Rent per month	Rent/home price per month as a percentage of monthly income	20 %, 30 %, 40 %, 50 %	Krueger et al. (2019); Patterson et al. (2017); Liao et al. (2015); Tu et al. (2016); Hoshino et al., 2011
	Commuting time	Commuting time to work by public transit	20 min or less, 20–40 min, 40–60 min, 60 min or more	
	Location	The location of the house	City centre, suburb	
	Share	Whether a rented property is shared with others	Share, no share	
BE attributes and neighbourhood attributes	Distance to public transportation	Distance from home to public transit	Less than 500m, 500m–1000m, 1000–1500m, 1500m or more	Krueger et al. (2019); Patterson et al. (2017); Liao et al. (2015)
BE attributes and neighbourhood attributes	Distance to essential amenities	Distance from home to destinations such as shops, restaurants, public libraries and schools	Walking distance, Less than 3 km, 3–10 km, 10 km or more	Krueger et al. (2019); Patterson et al. (2017); Liao et al. (2015)
	Distance to open space and green spaces	Distance from home to parks or forests	Less than 2 km, 2–3 km, 3–4 km, 4 km or more	
	Population density	The number of residents per square kilometre	5000 or less, 5000–10000, 10000–15000, 15000 or more	http://tjj.beijing.gov.cn/

Data availability

The data that has been used is confidential.

References

Ando, H., Ikegami, K., Nagata, T., Tateishi, S., Eguchi, H., Tsuji, M., et al., 2021. Effect of commuting on the risk of COVID-19 and COVID-19-induced anxiety in Japan, December 2020. Arch. Public Health 79, 1–10.

Axhausen, K.W., Hess, S., König, A., Abay, G., Bates, J., Bierlaire, M., 2006. State of the art estimates of the Swiss value of travel time savings. Arbeitsberichte Verkehrs-und Raumplanung 383.

Balbontin, C., Ortúzar, J.D.D., Swait, J.D., 2015. A joint best–worst scaling and stated choice model considering observed and unobserved heterogeneity: an application to residential location choice. Journal of choice modelling 16, 1–14.

Beck, M.J., Hess, S., Cabral, M.O., Dubernet, I., 2017. Valuing travel time savings: a case of short-term or long term choices? Transport. Res. E Logist. Transport. Rev. 100, 133–143.

Ben-Akiva, M., Morikawa, T., 1990. Estimation of travel demand models from multiple data sources. Transportation and traffic theory. In: Proceedings of the Eleventh International Symposium, Held July, pp. 18–20.

Berdejo-Espinola, V., Suárez-Castro, A.F., Amano, T., Fielding, K.S., Oh, R.R.Y., Fuller, R. A., 2021. Urban green space use during a time of stress: a case study during the COVID-19 pandemic in Brisbane, Australia. People and Nature 3 (3), 597–609.

Bhat, C.R., Guo, J.Y., 2007. A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. Transp. Res. Part B Methodol. 41 (5), 506–526.

Borgers, A., Timmermans, H., 1993. Transport facilities and residential choice behavior: a model of multi-person choice processes. Pap. Reg. Sci. 72 (1), 45–61.

Börjesson, M., 2008. Joint RP–SP data in a mixed logit analysis of trip timing decisions. Transport. Res. E Logist. Transport. Rev. 44 (6), 1025–1038.

Buckell, J., Hess, S., 2019. Stubbing out hypothetical bias: improving tobacco market predictions by combining stated and revealed preference data. J. Health Econ. 65, 93–102.
Bwambale, A., Choudhury, C.F., Hess, S., 2019. Modelling trip generation using Mobile phone data: a latent demographics approach. J. Transport Geogr. 76, 276–286.
Chen, R., Zhang, M., Zhou, J., 2023. Jobs-housing relationships before and amid COVID-19: an excess-commuting approach. J. Transport Geogr. 106, 103507.
Christidis, P., Cawood, E.N., Fiorello, D., 2022. Challenges for urban transport policy after the Covid-19 pandemic: main findings from a survey in 20 European cities. Transp. Policy 129, 105–116.
Conway, M.W., Salon, D., da Silva, D.C., Mirtich, L., 2020. How will the COVID-19 pandemic affect the future of urban life? Early evidence from highly-educated respondents in the United States. Urban Sci. 4 (4), 50.
Currie, G., Jain, T., Aston, L., 2021. Evidence of a post-COVID change in travel Behaviour–self-Reported expectations of commuting in Melbourne. Transport. Res. Pol. Pract. 153, 218–234.
de Palma, A., Vosough, S., Liao, F., 2022. An overview of effects of COVID-19 on mobility and lifestyle: 18 months since the outbreak. Transport. Res. Pol. Pract. 159, 372–397.
Delventhal, M.J., Kwon, E., Parkhomenko, A., 2022. JUE insight: how do cities change when we work from home? J. Urban Econ. 127, 103331.
Dryhurst, S., Schneider, C.R., Kerr, J., Freeman, A.L., Recchia, G., Van Der Bles, A.M., et al., 2022. Risk perceptions of COVID-19 around the world. In: COVID-19. Routledge, pp. 162–174.
Earnhart, D., 2002. Combining revealed and stated data to examine housing decisions using discrete choice analysis. J. Urban Econ. 51 (1), 143–169.
Eluru, N., Sener, I.N., Bhat, C.R., Pendyala, R.M., Axhausen, K.W., 2009. Understanding residential mobility: joint model of the reason for residential relocation and stay duration. Transp. Res. Rec. 2133 (1), 64–74.
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. Transport Geogr. 59, 146–155.
Ghosh, A., Nundy, S., Ghosh, S., Mallick, T.K., 2020. Study of COVID-19 pandemic in London (UK) from urban context. Cities 106, 102928.
Guevara, C.A., Ben-Akiva, M.E., 2013. Sampling of alternatives in logit mixture models. Transp. Res. Part B Methodol. 58, 185–198.
Guzman, L.A., Arellana, J., Cantillo-García, V., Ortúzar, J.D.D., 2021. Revisiting the benefits of combining data of a different nature: strategic forecasting of new mode alternatives. J. Adv. Transp. 2021 (1), 6672961.
Haque, M.B., Choudhury, C., Hess, S., dit Sourd, R.C., 2019. Modelling residential mobility decision and its impact on car ownership and travel mode. Travel behaviour and society 17, 104–119.
Harrington, D.M., Hadjiconstantinou, M., 2022. Changes in commuting behaviours in response to the COVID-19 pandemic in the UK. J. Transport Health 24, 101313.
Helveston, J.P., Feit, E.M., Michalek, J.J., 2018. Pooling stated and revealed preference data in the presence of RP endogeneity. Transp. Res. Part B Methodol. 109, 70–89.
Hensher, D.A., 1994. Stated preference analysis of travel choices: the state of practice. Transportation 21, 107–133.
Hensher, D.A., 2010. Hypothetical bias, choice experiments and willingness to pay. Transp. Res. Part B Methodol. 44 (6), 735–752.
Hensher, D.A., 2012. Accounting for scale heterogeneity within and between pooled data sources. Transport. Res. Pol. Pract. 46 (3), 480–486.
Hensher, D.A., Beck, M.J., Wei, E., 2021. Working from home and its implications for strategic transport modelling based on the early days of the COVID-19 pandemic. Transport. Res. Pol. Pract. 148, 64–78.
Hensher, D., Louviere, J., Swait, J., 1998. Combining sources of preference data. J. Econom. 89 (1–2), 197–221.
Hess, S., Lancsar, E., Mariel, P., Meyerhoff, J., Song, F., van den Broek-Altenburg, E., Alaba, O.A., Amaris, G., Arellana, J., Basso, L.J., Benson, J., Bravo-Moncayo, L., Chanel, O., Choi, S., Crastes dit Sourd, R., Cybis, H.B., Dorner, Z., Falco, P., Garzón-Pérez, L., Glass, K., Guzman, L.A., Huang, Z., Huynh, E., Kim, B., Konstantinus, A., Konstantinus, I., Larranaga, A.M., Longo, A., Loo, B.P.Y., Oehlmann, M., O'Neill, V., Ortúzar, J. de D., Sanz, M.J., Sarmiento, O.L., Moyo, H.T., Tucker, S., Wang, Y., Wang, Y., Webb, E.J.D., Zhang, J., Zuidgeest, M.H.P., 2022. The path towards herd immunity: predicting COVID-19 vaccination uptake through results from a stated choice study across six continents. Soc. Sci. Med. 298, 114800.
Hess, S., Palma, D., 2019. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. Journal of choice modelling 32, 100170.
Hoshino, T., 2011. Estimation and analysis of preference heterogeneity in residential choice behaviour. Urban Stud. 48 (2), 363–382.
Huang, J., Levinson, D., Wang, J., Jin, H., 2019. Job-worker spatial dynamics in beijing: insights from smart card data. Cities 86, 83–93.
Huang, J., Levinson, D., Wang, J., Zhou, J., Wang, Z.J., 2018. Tracking job and housing dynamics with smartcard data. Proc. Natl. Acad. Sci. 115 (50), 12710–12715.
Kim, J.H., 2006. Amenity valuing differentiation in residential location choice among income groups: a stated preference approach. International Journal of Urban Sciences 10 (1), 41–57.
Kleinhans, R., Kearns, A., 2013. Neighbourhood restructuring and residential relocation: towards a balanced perspective on relocation processes and outcomes. Hous. Stud. 28 (2), 163–176.
Krizek, K.J., 2003. Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? J. Am. Plann. Assoc. 69 (3), 265–281.
Krueger, R., Rashidi, T.H., Dixit, V.V., 2019. Autonomous driving and residential location preferences: evidence from a stated choice survey. Transport. Res. C Emerg. Technol. 108, 255–268.
Lee, B.K., Sohn, S.Y., Yang, S., 2014. Design guidelines for the dashilar, beijing open green space redevelopment project. Urban For. Urban Green. 13 (2), 385–396.
Liao, F.H., Farber, S., Ewing, R., 2015. Compact development and preference heterogeneity in residential location choice behaviour: a latent class analysis. Urban Stud. 52 (2), 314–337.
McFadden, D., 1972. Conditional Logit Analysis of Qualitative Choice Behavior. Frontiers in Econometrics. Academic Press, New York.
Patterson, Z., Darbani, J.M., Rezaei, A., Zacharias, J., Yazdizadeh, A., 2017. Comparing text-only and virtual reality discrete choice experiments of neighbourhood choice. Landsc. Urban Plann. 157, 63–74.
Qian, Y., Fan, W., 2020. Who loses income during the COVID-19 outbreak? Evidence from China. Res. Soc. Stratif. Mobil. 68, 100522.
Robbennolt, D., Haddad, A.J., Mondal, A., Bhat, C.R., 2023. Housing Choice in an Evolving Remote Work Landscape. The University of Texas at Austin. Technical paper, Department of Civil, Architectural and Environmental Engineering.
Shakib, S., Hawkins, J., Nurul Habib, K., 2022. Investigating the effects of COVID-19 pandemic on the perception of residential accessibility in greater Toronto and Hamilton area. Transp. Res. Rec., 03611981221086640
Stawarz, N., Rosenbaum-Feldbrügge, M., Sander, N., Sulak, H., Knobloch, V., 2022. The impact of the COVID-19 pandemic on internal migration in Germany: a descriptive analysis. Popul. Space Place 28 (6), e2566.
Stokenberga, A., 2019. How family networks drive residential location choices: evidence from a stated preference field experiment in Bogotá, Colombia. Urban Stud. 56 (2), 368–384.
Tao, Y., Petrović, A., van Ham, M., Fu, X., 2023. Residential relocation as a key event in commuting mode shift. Transport. Res. Transport Environ. 119, 103772.
Tayyaran, M.R., Khan, A.M., 2007. Telecommuting and residential location decisions: combined stated and revealed preferences model. Can. J. Civ. Eng. 34 (10), 1324–1333.
Thomas, M.J., Stillwell, J.C., Gould, M.I., 2016. Modelling the duration of residence and plans for future residential relocation: a multilevel analysis. Trans. Inst. Br. Geogr. 41 (3), 297–312.
Tillema, T., Van Wee, B., Ettema, D., 2010. The influence of (toll-related) travel costs in residential location decisions of households: a stated choice approach. Transport. Res. Pol. Pract. 44 (10), 785–796.
Tran, M.T., Zhang, J., Fujiwara, A., 2017. Interdependences between current choices and future expectations in the context of hanoians' residential location choices. Transport. Bus.: Transport Dynamics 5 (1), 38–55.
Tsoleridis, P., Choudhury, C.F., Hess, S., 2022. Utilising activity space concepts to sampling of alternatives for mode and destination choice modelling of discretionary activities. Journal of choice modelling 42, 100336.
Tu, G., Abildtrup, J., Garcia, S., 2016. Preferences for urban green spaces and peri-urban forests: an analysis of stated residential choices. Landsc. Urban Plann. 148, 120–131.
Van Acker, V., Witlox, F., 2009. Why land use patterns affect travel behaviour (or not). Toward a "state-of-the-art" conceptual framework and an appropriate modelling technique. Belgeo. Revue belge de géographie (1), 5–26.
Venter, Z.S., Barton, D.N., Gundersen, V., Figari, H., Nowell, M., 2020. Urban nature in a time of crisis: recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. Environ. Res. Lett. 15 (10), 104075.
Wang, Y., Wang, Y., Choudhury, C., 2020. Modelling heterogeneity in behavioral response to peak-avoidance policy utilizing naturalistic data of beijing subway travelers. Transport. Res. F Traffic Psychol. Behav. 73, 92–106.
Wang, Y., Choudhury, C., Hancock, T.O., Wang, Y., de Dios Ortúzar, J., 2024. Influence of perceived risk on travel mode choice during Covid-19. Transp. Policy 148, 181–191.
Whitehead, J.C., Lew, D.K., 2020. Estimating recreation benefits through joint estimation of revealed and stated preference discrete choice data. Empir. Econ. 58 (4), 2009–2029.
Wiig, A., Mahmud, M., Kolstad, I., Lujala, P., Bezu, S., 2020. Preparing to leave? Household mobility decisions in climate affected areas of coastal Bangladesh. CMI Working Paper WP 2020:4. Chr. Michelsen Institute.
Wolday, F., Böcker, L., 2023. Exploring changes in residential preference during COVID-19: implications to contemporary urban planning. Environ. Plan. B Urban Anal. City Sci. 50 (5), 1280–1297.
Yıldırım, M., Geçer, E., Akgül, Ö., 2021. The impacts of vulnerability, perceived risk, and fear on preventive behaviours against COVID-19. Psychol. Health Med. 26 (1), 35–43.
Zarrabi, M., Yazdanfar, S.A., Hosseini, S.B., 2021. COVID-19 and healthy home preferences: the case of apartment residents in Tehran. J. Build. Eng. 35, 102021.
Zhao, P., Gao, Y., 2023. Discovering the long-term effects of COVID-19 on jobs–housing relocation. Humanit. Soc. Sci. Commun. 10 (1), 1–17.
Zhao, Y., Pawlak, J., Sivakumar, A., 2022. Theory for socio-demographic enrichment performance using the inverse discrete choice modelling approach. Transp. Res. Part B Methodol. 155, 101–134.
Zolfaghari, A., Sivakumar, A., Polak, J.W., 2012. Choice set pruning in residential location choice modelling: a comparison of sampling and choice set generation approaches in greater London. Transport. Plann. Technol. 35 (1), 87–106.