Applied Geography 186 (2026) 103809

Contents lists available at ScienceDirect

Applied Geography

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

Does e-shopping weaken the link between land use and neighborhood shopping behavior? Evidence from large-scale mobile phone data

Fangyi Ding [a], Zhan Zhao [a,b,c,*], Yun Han [a], Yang Zhou [a], Yang Xu [d]

[a] Department of Urban Planning and Design, The University of Hong Kong, Hong Kong Special Administrative Region of China
[b] Urban Systems Institute, The University of Hong Kong, Hong Kong Special Administrative Region of China
[c] Musketeers Foundation Institute of Data Science, The University of Hong Kong, Hong Kong Special Administrative Region of China
[d] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China

ARTICLE INFO

Handling Editor: Neil Reid

Keywords:
Online shopping
Land use
Mobile signaling data
Travel distance
Travel frequency

ABSTRACT

Proximity-based land use planning has traditionally been viewed as a key strategy to manage travel demand and promote sustainable mobility. Yet, the rise of online activities, particularly the spread of e-shopping, may alter this rationale by reducing the need for physical proximity to consumption opportunities and potentially reshaping residents' interactions with neighborhood facilities. In this study, we draw on a one-month mobile signaling dataset from Shanghai, China, to provide large-scale behavioral evidence on this issue. Leveraging a Spatial Error Model (SEM), we examine whether e-shopping attenuates the influence of neighborhood land use features on shopping travel behaviors. Our findings show that residents in neighborhoods with higher levels of e-shopping engagement tend to make fewer offline shopping trips and travel shorter distances. More importantly, e-shopping moderates — and in many cases attenuates — the effects of key proximity-based features, including proximity to commercial centers, geographical location, road density, and transit accessibility. We also find significant heterogeneity across neighborhood types: communities with a higher share of elderly residents show greater sensitivity to these moderating effects, particularly in terms of shopping trip frequency. Overall, the results suggest that e-shopping may reshape the foundations of proximity-based planning. Our study further demonstrates the value of mobile signaling data in capturing these emerging dynamics, offering new insights for modeling travel demand and informing land use policies in the digital era.

1. Introduction

Rapid urbanization has exacerbated traffic congestion and environmental pollution (Bibri et al., 2020). Land use policies have conventionally been regarded as an effective strategy to reduce travel demand and alleviate urban congestion (Shi et al., 2020). In this context, proximity-based urban planning concepts — such as the compact city (Bibri et al., 2020), new urbanism (Congress for the New Urbanism, 2024), and the more recent "15-min city" (Moreno et al., 2021) — have emerged as leading paradigms of sustainable urbanism. These approaches emphasize the intensification of development and activities, encourage mixed land use and social diversity, highlight the role of spatial proximity in reducing the amount and distance of travel, and promote active modes such as walking and cycling.

As a form of non-compulsory travel, daily consumer activities like shopping and dining are largely constrained by spatial factors such as the location of consumption facilities and the accessibility of transport systems (Boarnet et al., 2011; Jiao et al., 2016; Xi et al., 2017). In recent years, however, the rapid expansion of e-shopping, which refers to business-to-consumer (B2C) purchases made through online channels (Le et al., 2022), has shown the potential to attenuate these spatial constraints (Shi et al., 2020). By fragmenting the shopping process (e.g., information gathering, evaluation, purchase, and delivery) and enabling it to be conducted remotely, e-shopping can effectively decouple consumption activities from neighborhood accessibility (Le et al., 2022), thereby modifying travel patterns (Ding & Lu, 2017; Ferrell, 2005; Shi, Cheng et al., 2021).

This raises questions about whether the proximity-based rationale of neighborhood planning remains effective in the e-shopping era. Since residents can now shop online even when local retail opportunities are limited, the importance of neighborhood density, diversity, and accessibility for shaping travel patterns may be reduced. As a result, differences in shopping travel behavior between retail-rich and retail-poor neighborhoods may become less pronounced. This shift could

* Correspondence to: Rm 809, 8/F, Knowles Building, The University of Hong Kong, Pokfulam Road, Hong Kong Special Administrative Region of China.
E-mail addresses: fyding@connect.hku.hk (F. Ding), zhanzhao@hku.hk (Z. Zhao), u3009502@connect.hku.hk (Y. Han), zhouyhk@hku.hk (Y. Zhou), yang.ls.xu@polyu.edu.hk (Y. Xu).

https://doi.org/10.1016/j.apgeog.2025.103809
Received 12 June 2025; Received in revised form 9 September 2025; Accepted 12 October 2025
Available online 23 October 2025
0143-6228/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

potentially diminish the effectiveness of proximity-based planning and design policies. Against this backdrop, it is of paramount importance to re-examine the impact of land use on travel behavior in an era of pervasive e-shopping. Consequently, we hypothesize that e-shopping loosens neighborhood-scale spatial constraints and attenuates the traditionally strong connections between neighborhood land use and daily shopping behavior (e.g., shopping travel distance and frequency).

Previous studies have largely examined the spatial diffusion of e-shopping adoption, investigating whether it first emerges in metropolitan areas as part of an innovation–diffusion process or in rural areas as an efficiency response to limited retail accessibility (Anderson et al., 2003; Beckers et al., 2018; Clarke et al., 2015; Farag et al., 2006). In addition, Shi et al. (2023) investigates the indirect effects of the spatial attributes on e-shopping through individuals' e-shopping attitudes. Although no clear consensus has been reached, these studies consistently demonstrate that e-shopping adoption is closely linked to spatial accessibility and land use context.

Building on this foundation, our study shifts the focus from the determinants of adoption to the consequences of e-shopping for travel behavior. Specifically, we ask whether the growing prevalence of e-shopping attenuates the extent to which neighborhood land use characteristics shape in-store shopping travel. Only a limited number of studies have explored this dimension at the individual scale (Shao et al., 2023; Shi et al., 2020).

In addition, a large body of research has explored the impact of e-shopping on shopping-related travel (Aldo et al., 2023; Arranz-López & Soria-Lara, 2022; Colaco & de Abreu e Silva, 2021; Ding & Lu, 2017; Ferrell, 2004). Our study differs in two important respects. First, we focus on the neighborhood scale rather than the individual or household level. To our knowledge, there has been limited attention to the influence of e-shopping within the context of proximity-based neighborhood planning. Second, we explore the potential of mobile signaling data in this field. While most previous studies rely on survey data—which, despite providing detailed personal attributes and accurate travel diaries, are susceptible to recall bias regarding e-shopping usage and travel behavior changes (Rotem-Mindali & Weltevreden, 2013; Shi, Cheng et al., 2021). In contrast, mobile signaling data offer an automated means of collecting quantitative information, such as monthly e-shopping duration. This reduces subjectivity inherent in survey responses and can lower both costs and respondent burden (Vij & Shankari, 2015). Furthermore, smartphone-enabled trajectory data provide rich spatial and temporal information for large populations (Oosterlinck et al., 2017), enabling a more comprehensive understanding of travel and activity behavior at the neighborhood level.

To bridge these gaps, our study complements existing research by examining the effects of land use on neighborhood-scale shopping travel distance and frequency under the influence of e-shopping. To obtain more refined and diverse land use features while mitigating potential reporting biases, we employ large-scale mobile signaling data from Shanghai, China. Our research seeks to answer the following key questions:

(1) How does e-shopping behavior influence offline shopping travel distance and frequency at the neighborhood scale?
(2) Do e-shopping activities moderate — particularly attenuate — the impacts of land use on travel patterns?
(3) If so, do these moderating effects vary across different socio-demographic groups?

In this study, we construct regression models to explore the relationship between e-shopping usage and shopping behavior (RQ1). We then incorporate interaction terms between e-shopping usage and land use factors to examine the moderating effects (RQ2). Finally, we estimate separate models across different socio-demographic groups to investigate the heterogeneity of these moderating effects (RQ3).

The remainder of the paper is organized as follows. Section 2 reviews the literature on the interplay of land use, e-shopping, and travel, as well as the potential of mobile signaling data. Section 3 presents the methodology, including the conceptual model, data, variables, modeling approach, and robustness analysis. Section 4 presents the empirical results addressing the three research questions and tests their robustness. Section 5 summarizes the key findings, and the final section highlights policy implications and limitation discussion.

2. Literature review

2.1. The interplay of land use features, E-shopping and travel

Land use factors in this study refer not only to land use functions such as residential, commercial, and retail, but also to geographic location attributes, and the built environment. Land use features have been demonstrated to play a crucial role in influencing both the distance and frequency of travel (Cervero & Kockelman, 1997; Ewing & Cervero, 2001). Specifically for shopping trips, residents in areas with low shopping accessibility or in exurban neighborhoods with low population density may make fewer in-store purchases (Boarnet et al., 2011; Farag et al., 2007; Li & Kim, 2020; Xi et al., 2017) and typically travel longer distance (Bönisch et al., 2021; Jiao et al., 2016). Conversely, residing in areas with more commercial land use correlates with shorter travel distance (Boarnet & Crane, 2001; Boarnet et al., 2011).

A related line of research has investigated the spatial diffusion of e-shopping adoption. Anderson et al. (2003) proposed two hypotheses regarding the effects of geographic context: the efficiency hypothesis, which predicts faster adoption in remote areas where e-shopping improves retail accessibility, and the innovation–diffusion hypothesis, which suggests that early adoption is concentrated in metropolitan areas before spreading to other regions. Following this framework, many empirical studies have shown that the two hypotheses are not mutually exclusive. Some researchers support the innovation–diffusion hypothesis (Clarke et al., 2015; Farag et al., 2006; Zhen et al., 2018), arguing that urban residents possess superior Internet access and technological proficiency compared to rural residents. Others provide evidence for the efficiency hypothesis, emphasizing accessibility issues as a major driver of e-shopping adoption (Cao et al., 2013; Loo & Wang, 2018; Ren & Kwan, 2009). In addition, Maat and Konings (2018) and Zhen et al. (2018) found that the impacts vary depending on product type. Beckers et al. (2018) showed that the urbanization level of an area does not significantly affect e-shopping probability, while Colaço and e Silva (2022) explored the geography of e-shopping in Lisbon and concluded that adoption can be linked to residential location preferences.

The potential effects of e-shopping on in-store shopping have also attracted considerable research attention, particularly with reference to the concepts of substitution (Ferrell, 2005; Shi et al., 2019; Shi, Shao et al., 2021), complementarity (Colaco & de Abreu e Silva, 2021; Farag et al., 2005; Shi, Cheng et al., 2021; Wei et al., 2024), modification (Ding & Lu, 2017; Ferrell, 2005; Shi, Cheng et al., 2021), and neutrality (Calderwood & Freathy, 2014; Sim & Koi, 2002). Among these studies, land use factors are often incorporated as control variables, such as shopping accessibility (Colaco & de Abreu e Silva, 2021; Ding & Lu, 2017; Ferrell, 2004; Lee et al., 2017) or geographical location (Farag et al., 2005; Shi et al., 2019; Zhou & Wang, 2014). More recent work has started to integrate richer land use attributes into analyses of the relationship between e-shopping and shopping travel (Aldo et al., 2023; Etminani-Ghasrodashti & Hamidi, 2020). For example, Aldo et al. (2023) examined the impact of e-shopping on shopping trips while controlling for factors such as rural/urban location, residential quality, facility supply, and public transport quality. They found that land use characteristics did not significantly influence the frequency of shopping trips. Despite these mixed findings of e-shopping and in-store shopping association, disadvantaged groups (e.g., low-income populations and the elderly) deserve particular attention. As slower adopters of e-shopping, they may be more vulnerable to reduced

shopping opportunities stemming from substitution effects, even when such effects are relatively small (Colaço & e Silva, 2022).

Some studies have suggested that online activities may attenuate the role of land use in shaping travel behavior (Elldér, 2017; Kwan et al., 2007). In the context of shopping behavior, Shi et al. (2020) investigated whether e-shopping for intangible services (i.e., services that are searched for and paid for online but consumed offline) attenuates spatial constraints on in-store shopping. They constructed two regression models for consumption via online versus offline channels and compared the coefficients of spatial attributes (urban vs. suburban). Their findings indicate that when services are purchased online, the negative association between spatial attributes and travel distance becomes weaker. Similarly, Shao et al. (2023) examined the relationship between transit accessibility and the space–time flexibility of shopping travel, as well as the moderating effects of online transit information searching. They constructed separate regression models for frequent and infrequent online users and compared the average marginal effects of transit accessibility. By further incorporating interaction terms between online searching and transit accessibility, they showed that significant negative marginal effects indicate that online searching largely attenuates the positive impact of transport accessibility on the space–time flexibility of shopping travel.

Nevertheless, these studies focus on specific shopping services (Shi et al., 2020) or relatively simple land use attributes such as geographic location (urban vs. suburban) (Shi et al., 2020) or transit accessibility (Shao et al., 2023). While at the individual level, personal attributes such as socio-economic characteristics or shopping attitudes may be more effective predictors of shopping behavior than land use attributes, at the community planning level, richer and more diverse land use factors need to be incorporated. The extent and mechanisms through which e-shopping moderates the effectiveness of proximity-based neighborhood planning at the neighborhood scale therefore warrant further investigation.

2.2. Data collection in E-shopping research and the potential of mobile phone-based data

Existing studies on e-shopping primarily rely on survey data (Calderwood & Freathy, 2014; Cao et al., 2010; Colaco & de Abreu e Silva, 2021; Ferrell, 2005; Shi et al., 2020; Sim & Koi, 2002). Survey-based data collection provides rich information on individual socio-demographics, user attitudes, and travel behavior, making it suitable for examining the relationship between e-shopping and in-store shopping behavior. However, one significant challenge lies in quantifying changes in travel behavior and e-shopping usage over time, which are susceptible to memory bias (Xiao et al., 2016) and social desirability bias (Cao, 2009), potentially leading to inaccuracies. A few studies have relied on activity diaries spanning one to three days (Farag et al., 2005, 2007; Zhou & Wang, 2014), but this period may be insufficient to capture stable impacts (Cao, 2009).

More recently, technological advances have enabled studies to track individual mobility using GPS devices. For instance, Ding and Lu (2017) combined GPS data with survey responses to investigate the relationship between online and in-store shopping behavior. Respondents carried GPS devices to record daily activities and trips, while surveys captured socio-demographic characteristics and Internet-use data. Conducted in one area of Beijing over seven days, the study obtained a sample of 709 individuals. Although this GPS-based method enables more accurate travel recording, the need for respondents to carry dedicated devices still creates inconvenience (Xiao et al., 2016).

In recent years, mobile signaling data has emerged as an automatic and effective approach to capturing users' spatiotemporal trajectories. When processed under stringent privacy regulations, aggregated mobile phone data ensures anonymity and legitimacy, making it well-suited for neighborhood-level studies (Zahnow et al., 2022). For example, Gao et al. (2023) examined grid-level activity patterns of rural residents during rapid urbanization in Chengdu, China. He and He (2024) and Zahnow et al. (2022) explored the relationship between neighborhood characteristics and walking behavior, while Bergroth et al. (2022) presented temporally dynamic population distributions in Helsinki's metropolitan area.

To date, however, few studies have explored the use of aggregated mobile phone data to examine how neighborhood-level e-shopping is associated with in-store shopping behavior and whether e-shopping moderates the relationship between neighborhood land use features and shopping trips. Utilizing aggregated and anonymized mobile phone data allows for large-scale, long-term tracking of population-level mobility across multiple neighborhoods, while avoiding the ethical concerns of individual-level GPS data and the respondent burden of surveys or travel diaries (Zahnow et al., 2022). These data provide detailed spatial and temporal information, making it possible to recover unreported trips (Xiao et al., 2016). In addition to mobility, mobile phone data may also contain information on app usage type and duration, which can be used to quantify e-shopping behavior and reduce recall bias.

Although mobile phone data complement survey data in several respects, they also have limitations. Unlike traditional travel surveys, passively generated mobile data often lack detailed socio-demographic attributes that are critical for travel behavior and policy analysis (Xu et al., 2015). Moreover, individuals' location records do not explicitly reveal travel context (e.g., purpose or mode) (Xu et al., 2015) and usually require additional rules or algorithms to interpret. Furthermore, due to the anonymity of big data and the scarcity of ground-truth validation data, it remains challenging to assess accuracy. As He and He (2024) note, research using mobile data must place greater emphasis on method validation, as analyses without proper verification may be of limited value.

Given these contexts, our study sets out to: (1) Complement existing research on e-shopping and in-store shopping behavior by leveraging long-term, large-scale travel records from mobile signaling data at the neighborhood level; (2) Provide new insights through a comprehensive investigation of how e-shopping moderates the impact of neighborhood land use features on shopping travel; and (3) Combine auxiliary data and robustness analysis to validate the use of mobile phone data for this purpose.

3. Data and methods

3.1. Conceptual framework

Fig. 1 presents the conceptual framework of this study. We focus on the interplay between e-shopping usage, land use, and in-store shopping, while controlling for neighborhood-level socio-demographic characteristics. Land use features (e.g., density, accessibility, and facilities) are assumed to directly influence shopping travel behavior, measured in terms of travel distance and frequency. E-shopping usage plays two potential roles within this framework. First, it may affect in-store shopping trips through substitution, complementarity, modification, or neutrality effects. Second, it may moderate — particularly attenuate — the influence of land use on shopping travel behavior. Overall, this model highlights the joint effects of land use and e-shopping in shaping travel patterns at the neighborhood scale.

3.2. Data source and preprocessing

In this study, a *shopping trip* or *consumption trip* is defined as a journey to a retail center or restaurant, irrespective of the establishment's size, business type, or whether a transaction or dining activity ultimately occurs. This inclusive definition aligns with previous research (Ramezani et al., 2021; Su et al., 2009). Additionally, *e-shopping* refers to the use of the internet for purposes such as browsing

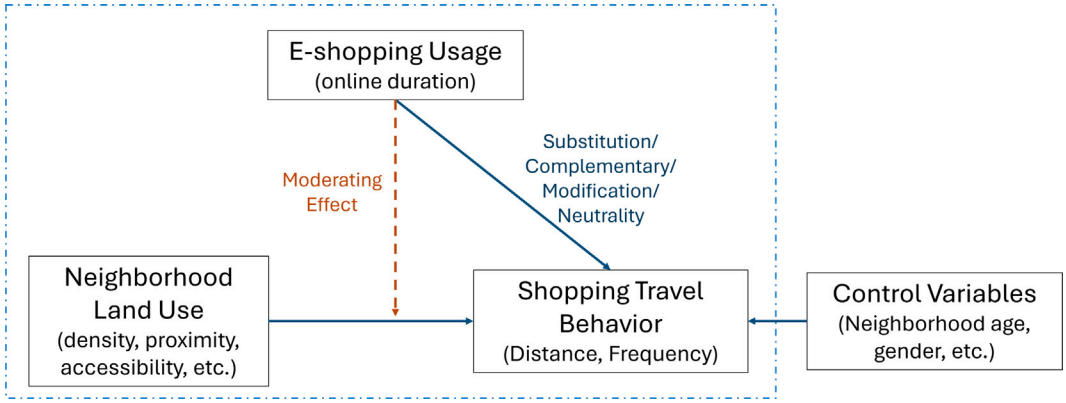


Fig. 1. Conceptual model of e-shopping, land use, and shopping travel behavior.

product information or making online purchases. This broad interpretation is consistent with prior studies that encompass a wide range of goods and services accessed digitally (Cao, 2009; Farag et al., 2007; Rotem-Mindali & Weltevreden, 2013).

This study utilizes one month of mobile signaling data from Shanghai in September 2020. During the data collection period, the COVID-19 pandemic was ongoing. However, according to statistics from the Shanghai Municipal Health Commission,[1] between March and September 2020 only two locally transmitted cases were reported, and no new local infections were recorded for three consecutive months leading up to September. The indigenous outbreak in Shanghai was therefore largely under control during this time. Moreover, the city had implemented a policy of precision prevention and control, which involved isolating and managing specific outbreak areas without significantly affecting normal life and travel in other regions. As a result, we can reasonably assume that no strict or mandatory mobility restrictions were imposed on residents within Shanghai during the study period. Nevertheless, given that lingering effects of the pandemic may still have reduced shopping trips and increased e-shopping usage (Colaço & de Abreu e Silva, 2024), we quantify the potential impact of this uncertainty, as described in Section 3.5.

The dataset is provided by China Unicom, one of the major mobile service providers in China. It contains user demographic details, online usage, and travel information. Demographic attributes cover variables such as gender and age. Internet usage is aggregated on a monthly basis and categorized by application type (e.g., shopping apps, music listening, mobile gaming). In this study, e-shopping activity is quantified based on the total time spent using shopping-related mobile applications.

A user's stay between two consecutive trips is considered as the completion of an activity. The basic mobile signaling data for each user consist of a sequence of stops, as shown in Table 1. Let the set of stops for user i be denoted as $S_i = \{s_k\}$, where s_k is the kth stop represented by $s_k = (st_k, et_k, lng_k, lat_k)$, including the start time, end time, longitude, and latitude. The duration of a stop is calculated as $t_k = et_k - st_k$. Stops that are too short to meet the minimum duration requirement are excluded as invalid. Specifically, valid stops must satisfy $t_k \geq \tau$, where the minimum stop duration τ is set to 30 min (Zhou et al., 2024).

Users' residential and work locations are inferred based on the cumulative dwelling time at each stop. To efficiently handle spatially adjacent stops and manage large-scale spatiotemporal data, the geographic area is divided into uniform 250m-by-250 m grid cells through geospatial gridding. The coordinates of stops are transformed into the corresponding grid indices, and each stop is updated to $s_k = (st_k, et_k, nx_k, ny_k)$, where nx_k and ny_k represent the horizontal and vertical grid indices.

Table 1
Data sample of user's trip records.

User ID	Start time	End time	Longitude	Latitude
328***126***	2020/9/1 0:15	2020/9/1 9:56	116.097512	39.959644
328***126***	2020/9/1 10:16	2020/9/1 11:34	116.212363	39.903882
328***126***	2020/9/1 11:52	2020/9/1 15:29	116.233285	39.899247

Residential and work locations (grids) are identified from the areas where users spend most of their time during night and daytime hours on weekdays, respectively. Specifically, the user's cumulative monthly dwelling time in each grid is computed. The grid with the highest dwelling time between 9:00 p.m. and 8:00 a.m. is designated as the residential location, while the one with the highest dwelling time between 9:00 a.m. and 5:00 p.m. is identified as the work location.

With users' residential and work locations identified, the next step is to infer shopping trips. The Shanghai travel survey (described in Appendix A) shows that 90.5% of shopping trips are home-based (i.e., originating from home). In addition, compared with workplaces, shopping activities are more frequently conducted in shops near home (Ding & Lu, 2017). Therefore, our study focuses on residential neighborhoods and only considers home-based shopping trips.

In this study, all trips are treated separately. For instance, if a shopping activity occurs during a commute from home to work, it is identified as two distinct trips: the first as a home-based shopping trip and the second as a non-home-based work trip. We employ a two-step procedure to identify home-based shopping trips. First, we exclude trips whose origins are not homes and those whose destinations are workplaces. This ensures that the remaining trips are home-based and non-commuting. Second, for these remaining trips, a POI–activity matching method is applied, where each trip destination is matched to its nearest POI (Point of Interest). Trips are classified as shopping trips if the matched POI belongs to commercial categories such as restaurants, retail stores, or markets. Although simple, this POI–activity matching method and its variants have been widely adopted and proven effective in trip purpose inference (Dhananjaya & Sivakumar, 2022; Meng et al., 2017; Xie et al., 2009). Its effectiveness is validated in the Appendix A through a comparison with travel survey data, using labeled trip purposes as ground truth.

To protect user privacy, all data are aggregated to the 250 m-by-250 m grid level for subsequent analysis. Since the focus is on home-based shopping trips, residential communities are delineated using Area of Interest (AOI) data. Grid cells that intersect with these AOIs are designated as residential neighborhoods, and China Unicom users residing in those grids constitute the study sample.

[1] https://wsjkw.sh.gov.cn/.

For data cleaning, we dropped grid cells with insufficient users (fewer than 10). As a result, 5353 grid cells are kept for analysis, which covers about 0.7 million users. As depicted in Fig. 2, the grid cells cover sections of Shanghai, including both the central city and its suburbs. From Fig. 2(a–b), it is evident that neighborhoods in the central city predominantly showcase characteristics of short-distance shopping trips, and neighborhoods situated in the central city and along its periphery register higher shopping trip frequencies. In contrast, neighborhoods on the outskirts of Shanghai display longer trip distances and fewer trip frequencies, implying a greater influence of spatial constraints on these areas.

To examine the interactions between e-shopping, urban contexts, and consumption trips, we use the shopping-related average travel distance per trip and monthly frequency of trips per capita in each grid cell as dependent variables. Influencing factors include the monthly shopping app usage time per resident, socio-demographic distribution, and land use features.

3.3. Variable selection

In this study, land use features include the geographical location (e.g., whether the grid cell located in the central city or not), density (e.g., residential density, job density), diversity (e.g., land use diversity), design (e.g., road network density), shopping opportunity (e.g., distance to nearest commercial centers, density of consumption facilities), and transit accessibility (e.g., the distance to nearest subway/bus station, subway/bus line count). The details of these attributes are as follows:

(1) ***Central city***. This variable distinguishes between grids within the central city and those outside it. According to the Shanghai Master Plan (Shanghai Municipal People's Government, 2018), the city center is defined as the area within the outer ring highway, covering approximately 664 square kilometers. The suburbs encompass areas beyond the central city boundaries. This division is illustrated in Fig. 2.

(2) ***Density***. The residential and job density per grid cell is inferred from mobile signaling data. Leveraging users' daily travel–activity information, we can deduce users' residential and work locations as detailed in Section 3.2. Taking into account users from other operators, we estimate the population residing or working in a grid cell through a weighted sum (individual weights are provided in the dataset). The residential/job density is finally calculated as

$$D = \frac{\sum_{i=1}^{N} w_i}{a} \tag{1}$$

where w_i represents the weight of individual i, N denotes the population of China Unicom users living or working in the grid cell, and a signifies the area of the grid cell.

The distributions of both residential and job density are displayed in Fig. 2(c–d). It is observed that the densities within the central city are higher than those in the surrounding areas. This concurrence with the real-world distribution validates that our data effectively captures a genuine representation of both places of work and residence.

(3) ***Diversity***. Land use diversity measures the mix of land uses in neighborhoods. It is quantified using the Shannon entropy of land use categories within a 500 m buffer surrounding the chosen grid cell. Lower entropy values indicate areas with minimal land use diversity, while higher values suggest a broader diversity of land use. To compute the Shannon entropy index, we employ the POI data as follows:

$$H = -\sum_{j=1}^{S} p_j \ln p_j \tag{2}$$

where p_j represents the percentage of the jth POI category, and S denotes the number of categories. The spatial distribution of land use diversity is showcased in Fig. 2(e).

(4) ***Shopping opportunity***. Shopping opportunities are assessed based on proximity to commercial centers and the density of shopping facilities. To determine these indices, we factor in POIs related to consumption activities, such as shopping, dining, and other everyday consumption establishments.

For gauging accessibility to commercial centers, we employ DBSCAN (density-based spatial clustering of applications with noise) (Ester et al., 1996) to group consumption POIs, thereby identifying prominent commercial hubs in Shanghai. These clusters are stratified into four tiers based on their sizes. Moving from tier 1 to tier 4, there is an increase in the number of clusters, while the POI density within each cluster decreases, transitioning from densely packed to more dispersed.

Fig. 2(f) showcases the four commercial center tiers. It is shown that tier 1 centers predominantly reside in the central city; tier 2 centers are found in both the central city and its outskirts, albeit with a sparse distribution; centers of tiers 3 and 4 have a broader spread across the city. These commercial tiers can be conceptualized as city level, district level, sub-district level, and local community level, respectively. The accessibility of a neighborhood to shopping centers is gauged by its distance to the closest commercial center within each respective tier.

Regarding the density of consumption establishments, it is evaluated within a 500 m buffer surrounding each neighborhood as:

$$D_s = \frac{N_s}{a_b} \tag{3}$$

where N_s denotes the count of consumption-related POIs within a 500 m buffer area. Fig. 2(g) delineates the distribution of such establishments. Notably, the central city and specific area cores exhibit a heightened facility density.

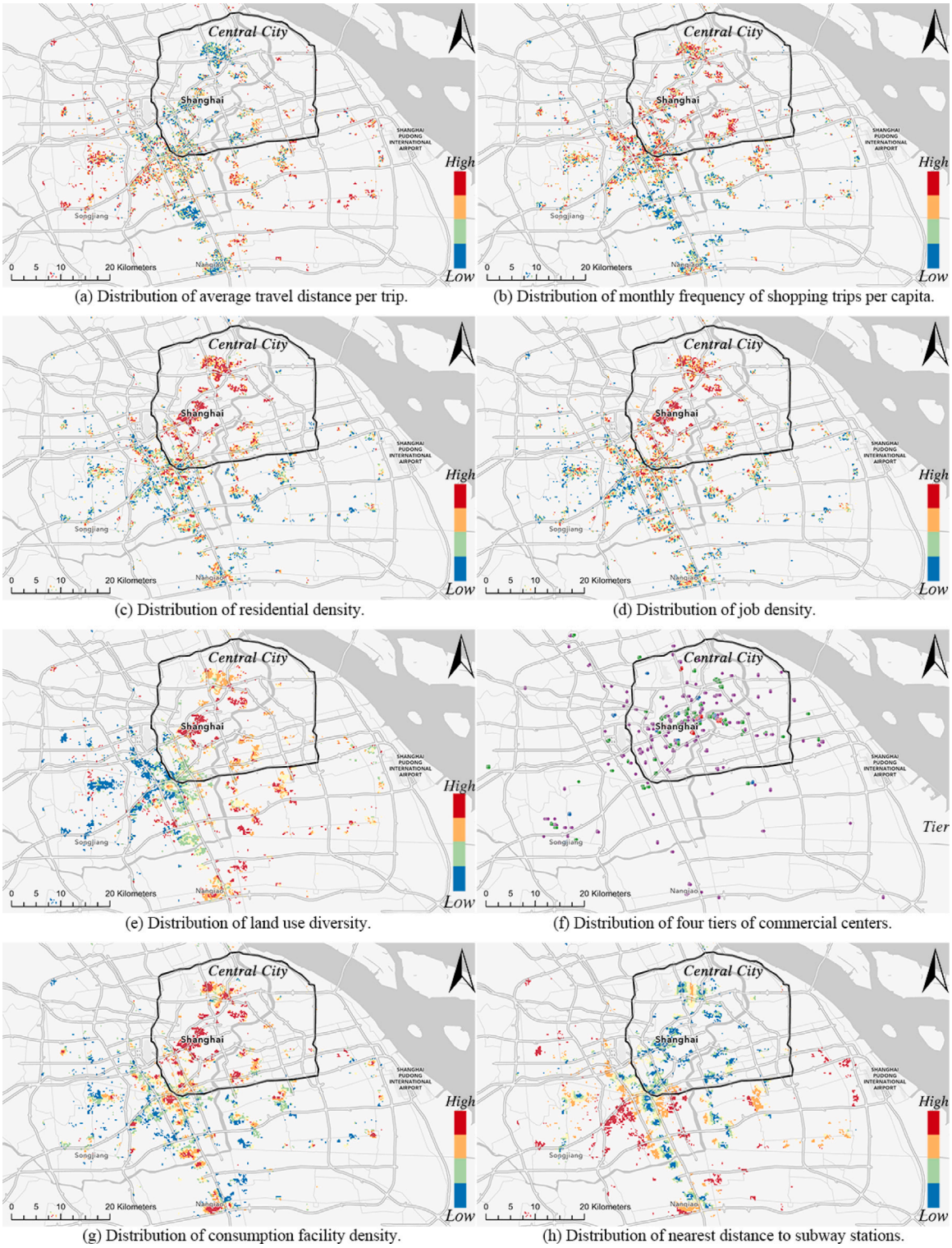
(5) ***Design***. The road network information is sourced from the OpenStreetMap (OSM). We categorize and compute the densities of roads into three distinct types: major roads, minor roads, and footways. Specifically, major roads encompass well-maintained highways and primary arterial routes that bridge cities and regions; minor roads are typically local streets serving specific neighborhoods or areas; footways are exclusively designed routes for pedestrians, facilitating walking and cycling. The road density for each category t is calculated by the length of roads per unit area within a 500 m buffer area:

$$D_r^t = \frac{\sum_{i=1}^{N_r^t} l_i^t}{a_b} \tag{4}$$

where l_i^t is the length of road segment i of type t in the grid cell, N_r^t is the number of road segments of type t in the area, and a_b signifies the area of the buffer.

(6) ***Transit accessibility***. Public transportation data is collected through web scraping. Transit proximity is indicated by the distance to the closest subway station or bus stop. Meanwhile, the number of subway or bus lines within a 500 m buffer of a grid cell represents transit connectivity. Fig. 2(h) illustrates the distribution of proximity to subway stations. It is clear that neighborhoods in the central city tend to be closer to subway stations compared to those in the suburbs.

Summary statistics of these variables for valid grid cells is presented in Table 2. It is evident that the monthly average time spent on online shopping per resident per grid cell is 33.5 h. In terms of population composition, the proportion of males slightly outweighs that of females, and the demographic distribution skews towards young and middle-aged individuals compared to children and the elderly.



(a) Distribution of average travel distance per trip.
(b) Distribution of monthly frequency of shopping trips per capita.
(c) Distribution of residential density.
(d) Distribution of job density.
(e) Distribution of land use diversity.
(f) Distribution of four tiers of commercial centers.
(g) Distribution of consumption facility density.
(h) Distribution of nearest distance to subway stations.

Fig. 2. Distribution of travel-related variables and certain land use features.

3.4. Model specification

To understand the influencing factors of consumer travel behavior, we regressed the consumption trip distance and frequency for different neighborhoods (i.e., grid cells) on their respective demographics, e-shopping usage, and land use features within a month. We initially employed a linear regression model, based on the Ordinary Least Squares (OLS) method, to test for correlation. The results revealed that most independent variables had a significant association with the dependent variables. However, urban mobility data may exhibit spatial dependence (Ding et al., 2025), making an OLS regression biased and subsequently leading to inaccurate results.

To address spatial autocorrelation, we turn to spatial regression models. Common spatial regression models include the Spatial Lag Model (SLM) and Spatial Error Model (SEM), the former models spatial dependency in the outcome while the latter in the residuals. To identify

Table 2
Descriptive analysis of the variables.

Variable	Mean	Std	Definition
ICT-related variables			
Avg. online time	33.55	17.04	Average time spent online shopping per person (hours)
Socio-demographic variables			
% of male	0.51	0.09	Male percentage
% of female	0.49	0.09	Female percentage
% <18 yrs old	0.13	0.06	Percentage of individuals under 18 years old
% 18–39 yrs old	0.49	0.09	Percentage of individuals aged 18–39 years
% 40–59 yrs old	0.22	0.08	Percentage of individuals aged 40–59 years
% >60 yrs old	0.16	0.08	Percentage of individuals above 60 years old
Avg. house price	27.65	14.77	Average house price (1000 yuan/m^2)
Land use features			
Residential density	11.98	16.62	Residential density per grid (1000 people/km^2)
Job density	12.55	18.20	Job density per grid (1000 people/km^2)
Land use diversity	0.27	0.01	Shannon entropy of land use within 500 m buffer of the grid cell
Dist. to nearest tier1	15.25	9.54	Distance to the closest commercial centers of tier 1 (km)
Dist. to nearest tier2	9.57	6.43	Distance to the closest commercial centers of tier 2 (km)
Dist. to nearest tier3	5.78	5.95	Distance to the closest commercial centers of tier 3 (km)
Dist. to nearest tier4	2.38	2.52	Distance to the closest commercial centers of tier 4 (km)
Consumption facility density	92.83	89.14	Density of consumption establishments within 500 m buffer (Count per km^2)
Central city[a]	0.31	0.46	Whether located in the central city (1 = yes, 0 = no)
Major road density	5.09	2.82	Major road density within 500 m buffer (km/km^2)
Minor road density	2.52	1.68	Minor road density within 500 m buffer (km/km^2)
Footway density	0.64	1.30	Footway density within 500 m buffer (km/km^2)
Dist. to nearest subway	1.97	2.61	Distance to the closest subway station (km)
Dist. to nearest bus	0.10	0.12	Distance to the closest bus station (km)
No. of subway lines	0.50	0.72	Number of subway lines within 500 m buffer
No. of bus lines	11.73	7.86	Number of bus lines within 500 m buffer

Note: The "Mean" column represents the percentage of instances where the variable equals 1, and the "Std" column represents the percentage of instances where the variable equals 0.
[a] For binary variables.

the most suitable model, we conduct diagnostic tests, including the Lagrange Multiplier tests for error dependence (LM error test) and spatial lag dependence (LM lag test). In the event of significant results for both tests, robust LM error and lag tests are conducted. Using a straightforward decision rule (Anselin, 1988; Anselin et al., 1996), we determine the source of dependence (lag or error) and select the most suitable model.

Specifically for our study, the SEM is utilized. The SEM assumes that spatial dependencies materialize through the error terms, rather than the dependent variable. The basic vector form of the SEM (LeSage & Pace, 2009) can be expressed as

$$y = X\beta + u \tag{5}$$

$$u = \lambda W u + \varepsilon \tag{6}$$

where y is an $n \times 1$ vector representing the dependent variable across n spatial units. X is an $n \times p$ matrix, where each element x_i denotes the values of the pth variable for the ith grid cell. β is a $p \times 1$ vector of coefficients to be estimated. W is an $n \times n$ matrix of spatial weights, where a spatial weight $w_{i,j}$ quantifies the geographical relationship between grid cell i and grid cell j. In this study, we utilize the Moore neighborhood, which defines a collection of neighboring grid cells that share at least one vertex, to denote the spatial adjacency relationship. λ is a scalar representing the spatial dependence in the sample observations, and ε is a vector of normally-distributed random errors.

3.5. Robustness analysis

Robustness analysis is an essential step in empirical research to evaluate whether estimated results are sensitive to potential data errors and alternative assumptions. The underlying principle is to examine the stability of key findings under controlled perturbations. A typical approach is to manually inject noise into class labels or attributes to simulate plausible levels of misclassification or measurement error (Zhu & Wu, 2004).

In this study, we conduct a perturbation-based robustness analysis under multiple scenarios, as summarized in Table 3. Specifically, we introduce controlled noise into (i) the dependent variables — average consumption trip distance and frequency — to simulate misclassification of shopping trips; and (ii) the online shopping time variable to account for uncertainty in translating online browsing/searching into actual purchases. In addition, to reflect behavioral shifts during the COVID-19 pandemic, where individuals may have reduced shopping travel distances and frequencies alongside increased online shopping, we inject positive noise into dependent variables and negative noise into online shopping time as normal period simulation. By comparing model estimates across varying perturbation levels, we assess the robustness of our results in terms of coefficient magnitudes, signs, and statistical significance.

Formally, let y_d and y_f denote the observed dependent variables at the neighborhood level, representing average shopping trip distance per trip and average monthly frequency of trips per capita, respectively. Due to potential misclassification of trip purposes, some non-shopping trips may be misclassified as shopping trips, while some shopping trips may be misclassified as non-shopping trips. This leads to biased observed values compared with the true y_d and y_f. We introduce a perturbation term ϵ at the neighborhood level, such that the simulated true variables are expressed as:

$$y'_d = y_d \cdot (1 + \epsilon), \tag{7}$$

$$y'_f = y_f \cdot (1 + \epsilon). \tag{8}$$

It is important to note that since trip classification errors are independent across trips, the errors tend to cancel out at the aggregate level. That is, if the individual trip-level classification error rate is p, then at the neighborhood level the aggregated deviation ϵ is much smaller than p. As shown in Appendix B, the standard deviation of ϵ at the neighborhood level can be expressed as:

$$\sigma_{\bar{\epsilon}} = \frac{\sqrt{pq/n}}{q}, \tag{9}$$

Table 3
Scenarios for perturbation-based robustness analysis.

Scenario	Explanation	Perturbed variables
(1) Dependent variable uncertainty	Misclassification in shopping trip recognition	Avg. dist, Avg. freq
(2) Independent variable uncertainty	Uncertainty in mapping online browsing to actual purchasing	Avg. online time
(3) Pandemic-related uncertainty	COVID-19 may alter travel and online consumption patterns	Avg. dist, Avg. freq, Avg. online time

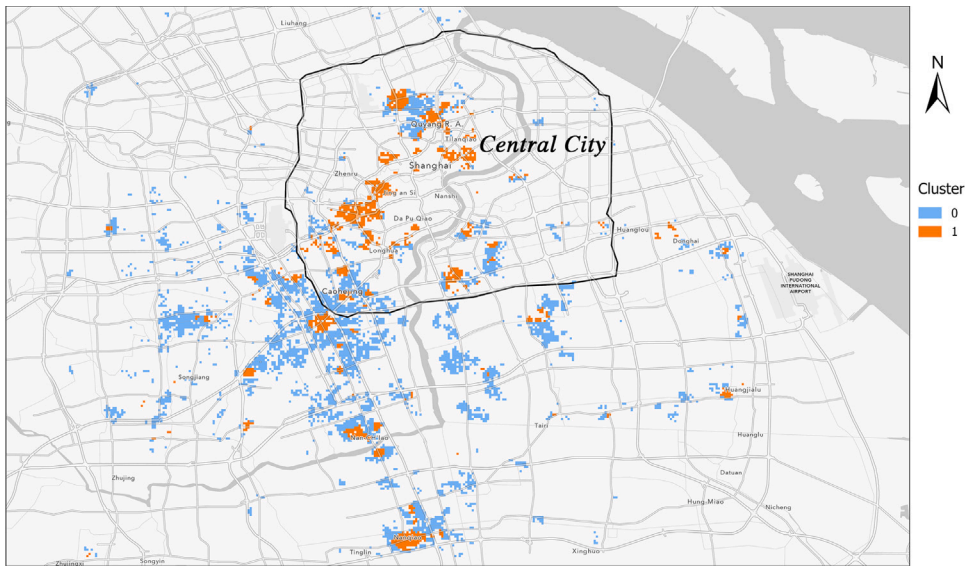


Fig. 3. The spatial distribution of two groups of neighborhoods.

where $q = 1 - p$ is the classification success rate, and n is the number of shopping trips in the neighborhood. For example, if $p = 0.2$ (i.e., 20% error rate as reported in Appendix A) and the average number of trips per neighborhood is $n = 200$, then the standard deviation becomes $\sigma_{\bar{\epsilon}} = \frac{\sqrt{0.2 \times 0.8/200}}{0.8} \approx 0.035$. This indicates that although the individual-level error rate is 20%, the uncertainty at the averaged neighborhood level is only about 3.5%, due to the law of large numbers.

Similarly, let x_t denote the observed online consumption usage time at the neighborhood level. We assume that actual online purchases are positively correlated with online shopping app usage time, with an average transfer rate α. Since α varies across users with noise ϵ, the simulated true online purchasing can be expressed as:

$$x'_t = x_t \cdot \alpha \cdot (1 + \epsilon). \tag{10}$$

Because α only scales the variable and does not affect model inference (e.g., coefficient signs or significance), we can simplify to:

$$x'_t = x_t \cdot (1 + \epsilon). \tag{11}$$

4. Results

4.1. E-shopping impact on travel behavior under homogeneous land use characteristics

Before delving into the relationship between online and offline shopping, it is necessary to test if varying levels of online usage significantly influence travel behavior. To control for the influence of land use, we cluster the sampled neighborhoods based on their land use features. For this, we employ the K-means clustering method. The Silhouette Coefficient aids in pinpointing the optimal number of clusters. The results suggest that classifying into two subgroups is the most effective. The spatial distribution of these neighborhoods and their corresponding land use feature distributions are presented in Figs. 3 and 4, respectively.

The spatial characteristics of the neighborhoods reveal that those in Group 1 are more centrally located than those in Group 0. Predominantly, Group 1 neighborhoods are situated in the central city and the core areas of various regions, boasting higher densities, superior transport accessibility, and more convenient consumer facilities. Hence, we label Group 0 as "peripheral neighborhoods" and Group 1 as "central neighborhoods".

To delve deeper, we further categorize these two primary groups into four subgroups based on e-shopping usage, aiming to analyze the discrepancies in travel behavior attributable to variances in online shopping dependences. Neighborhoods with e-shopping usage time below the average are designated as having a "high e-shopping level", while those exceeding this threshold are identified with a "low e-shopping level". Consequently, four distinct subsets integrating land use features and e-shopping usage patterns are generated.

To discern the potential disparities in travel behavior, we conduct a t-test for both average trip distance and monthly average travel frequency. The t-statistic quantifies the difference between the means of the two groups, whereas the p-value gauges the likelihood of observing the given difference, assuming the null hypothesis holds true. In this context, the null hypothesis posits that no significant difference exists between subgroups sharing similar land use characteristics but varying online usage levels. The outcomes of the t-test are detailed in Table 4. At a significance threshold of 0.1, all null hypotheses are refuted, suggesting that online shopping indeed influences travel behavior across neighborhoods significantly.

4.2. Relationship between online and offline shopping

Based on the insights from Section 4.1, we further explore the specific influence of e-shopping on travel behavior. Specifically, we aim to determine if online shopping has a substitution, complementary, modification, or neutral effect on travel patterns. Before proceeding with spatial regression, we perform a Variance Inflation Factor (VIF) diagnosis on all variables to examine any potential multicollinearity

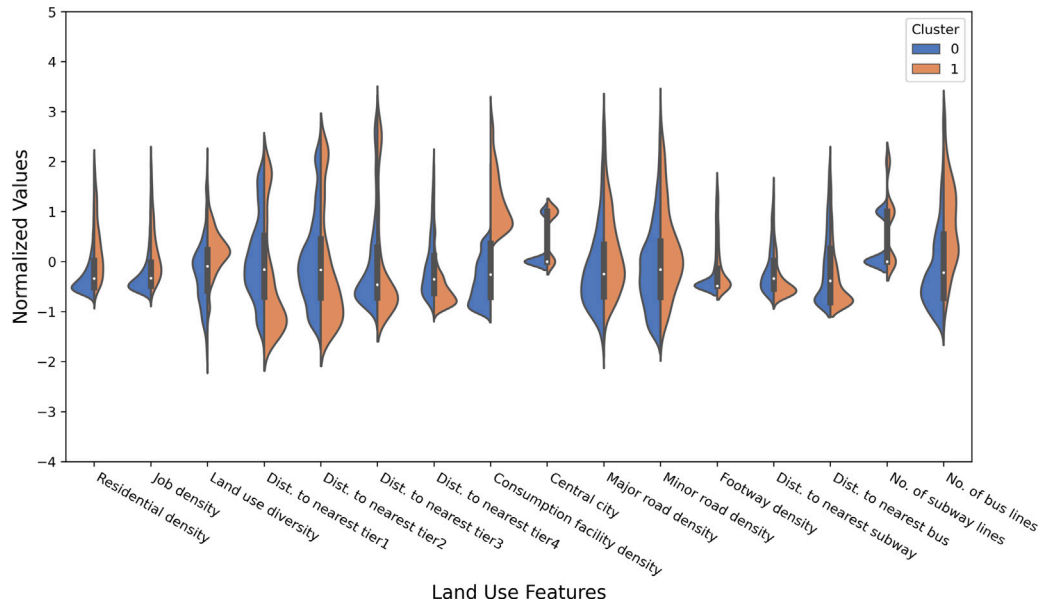


Fig. 4. Land use feature distribution of two groups of neighborhoods.

Table 4
t-test for average trip distance and monthly average travel frequency.

Variable	Subset 1	Subset 2	*t*-statistic
Avg. Dist	Central - high e-shopping level	Central - low e-shopping level	1.93 ($p \leq 0.1$)
Avg. Freq	Central - high e-shopping level	Central - low e-shopping level	7.06 ($p \leq 0.01$)
Avg. Dist	Peripheral - high e-shopping level	Peripheral - low e-shopping level	4.74 ($p \leq 0.01$)
Avg. Freq	Peripheral - high e-shopping level	Peripheral - low e-shopping level	13.46 ($p \leq 0.01$)

issues. It is found that the VIF values for all variables are below the threshold of 5, suggesting that significant multicollinearity is not a concern in our data.

Subsequently, we conduct spatial diagnostic tests to examine spatial dependence and determine the most suitable model. The results from these six statistical estimations, in relation to both shopping travel distance and frequency, are listed in Table 5. The Moran's I index in the regressions for both travel distance and frequency is found to be significant, which confirms the presence of spatial dependence in our data. When assessing the LM and robust LM tests, the error test results exhibit higher significance compared to the lag tests. This suggests that spatial error dependence is more prevalent in our scenario, making the SEM a more fitting choice for this study.

For enhanced model selection, we utilize information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Models with lower AIC or BIC values are typically regarded as superior, striking a balance between model fit and parsimony. Referring to Table 5, it is evident that both the AIC and BIC values for the SEM are lower than those for the SLM. This suggests that the SEM offers a more robust interpretation of the relationship between the explanatory variables and travel behavior compared to the SLM. Furthermore, since the SEM can accommodate potential explanatory variables that might have significant yet unaccounted associations with travel behavior, we have chosen to apply the SEM in our research.

In the base SEM, we evaluate the impact of neighborhoods' e-shopping usage on their travel behavior, with a specific focus on identifying substitution, complementary, modification, or neutrality relationships. To achieve better goodness-of-fit and obtain normal distributions, a logarithmic transformation is applied to the dependent variables, namely, the average distance traveled per trip and the monthly frequency of per capita shopping trips in each grid cell. Controlling for spatial dependence, the maximum likelihood method is employed to derive the SEM coefficients. The results for these two base models are presented in Table 6.

Table 5
Spatial diagnostic tests and fitness measures.

Test	Value for Dist	Value for Freq
Diagnostics for spatial dependence		
Moran's I (error)	18.320 ***	35.965***
Lagrange Multiplier (lag)	3.814*	985.055***
Robust LM (lag)	22.837***	60.698***
Lagrange Multiplier (error)	321.139***	1257.398***
Robust LM (error)	340.162***	333.041***
Lagrange Multiplier (SARMA)	343.976***	1318.096***
Measures of fit		
AIC (lag)	5667.544	13 919.748
AIC (error)	**5384.363**	**13 704.036**
BIC (lag)	5825.594	14 077.798
BIC (error)	**5535.828**	**13855.501**

Note: ***, ** and * represent the significance at the confidence level of 99%, 95% and 90%, respectively.

The e-shopping usage time coefficient is negative for both shopping trip distance and frequency, suggesting that residents in neighborhoods with higher e-shopping engagement tend to travel less frequently and cover shorter distances for offline shopping. This underscores a substitution relationship between online and offline shopping with respect to travel frequency. Concurrently, the results also reveal a modification effect on travel distance. This indicates that online shopping, which allows for remote evaluation and purchase of goods, can replace physical shopping trips, particularly longer ones, thereby reducing the average travel distance per trip.

Land use features can significantly influence consumer travel behavior. Specifically, residents in areas boasting diverse land uses often have longer shopping trips. One reason might be that diverse land uses can increase the likelihood of forming multi-activity travel chains, thus potentially increasing the shopping distance (Ding et al., 2017). There

Table 6
SEM results of base model and full model.

Variables	Avg. Dist #				Avg. Freq #			
	Base model		Full model		Base model		Full model	
	Coef.	Std	Coef.	Std	Coef.	Std	Coef.	Std
ICT-related variables								
Avg. online time	−1.72e−03***	3.75E−04	−9.49E−04	7.75E−03	−1.02e−02***	8.19E−04	2.26E−02	1.70E−02
Socio-demographics								
% of male	4.36E−02	6.16E−02	3.61E−02	6.12E−02	−9.23E−02	1.32E−01	−7.52E−02	1.31E−01
% <18 yrs old	1.80E−01	1.10E−01	1.32E−01	1.09E−01	−8.78E−02	2.37E−01	−9.06E−02	2.37E−01
% 18–39 yrs old	5.73E−02	8.11E−02	3.26E−02	8.05E−02	−2.42E−02	1.75E−01	−7.23E−03	1.74E−01
% >60 yrs old	−5.67E−02	1.07E−01	−7.34E−02	1.06E−01	−7.56E−02	2.27E−01	−4.46E−02	2.27E−01
Avg. house price	3.26E−04	6.29E−04	3.92E−04	6.18E−04	2.22E−04	1.51E−03	2.00E−04	1.50E−03
Land use features								
Residential density	3.91E−05	4.62E−04	5.81E−04	8.54E−04	4.68E−04	1.02E−03	−4.92E−04	1.84E−03
Job density	−1.06E−04	3.97E−04	−7.14E−04	8.12E−04	1.80E−04	8.51E−04	1.04E−03	1.73E−03
Land use diversity	1.69e+00***	5.45E−01	2.50e+00**	1.17E+00	−4.54e+00***	1.26E+00	−2.84E−01	2.59E+00
Dist. to nearest tier1	2.18e−02***	1.84E−03	1.77e−02***	3.13E−03	−1.69e−02***	4.51E−03	−2.68e−02***	7.03E−03
Dist. to nearest tier2	−2.10e−02***	2.30E−03	−1.54e−02***	4.10E−03	−5.98E−03	5.72E−03	1.93e−02**	9.35E−03
Dist. to nearest tier3	−5.06e−03**	2.56E−03	6.50E−03	4.51E−03	−1.08e−02*	6.36E−03	−2.89E−03	1.03E−02
Dist. to nearest tier4	2.34e−02***	4.31E−03	1.59e−02*	8.14E−03	8.70E−03	1.07E−02	−1.39E−02	1.84E−02
Consumption facility density	−5.35e−04***	1.17E−04	−1.22E−04	2.16E−04	5.26e−04*	2.92E−04	3.40E−04	4.88E−04
Central city	−2.75E−02	2.75E−02	−9.66e−02**	4.84E−02	1.16e−01*	6.85E−02	−1.25E−01	1.09E−01
Major road density	−3.11E−03	3.03E−03	−1.39e−02**	5.95E−03	1.36e−02*	7.49E−03	2.49e−02*	1.36E−02
Minor road density	1.85e−02***	4.59E−03	−2.32e−02***	8.53E−03	−5.40E−03	1.15E−02	2.64E−02	1.95E−02
Footway density	−6.23E−04	5.85E−03	2.20e−02**	9.84E−03	−1.71E−03	1.46E−02	−1.94E−02	2.24E−02
Dist. to nearest subway	9.91e−03**	4.72E−03	−5.19E−04	8.81E−03	−7.91E−03	1.17E−02	1.73E−02	2.00E−02
Dist. to nearest bus	2.82E−02	5.60E−02	1.99e−01*	1.18E−01	9.49E−02	1.26E−01	−2.46E−01	2.58E−01
No. of subway lines	1.42E−02	1.18E−02	1.86E−02	2.23E−02	3.94E−02	2.86E−02	1.02e−01**	4.98E−02
No. of bus lines	−2.23e−03*	1.30E−03	−2.33E−03	2.51E−03	1.38E−03	3.20E−03	−3.80E−03	5.64E−03
Interaction with Avg. online time								
× Residential density			−1.83E−05	2.54E−05			3.48E−05	5.40E−05
× Job density			2.37E−05	2.52E−05			−2.86E−05	5.35E−05
× Land use diversity			−1.94E−02	2.85E−02			−1.12e−01*	6.25E−02
× Dist. to nearest tier1			2.05e−04**	9.90E−05			4.03e−04*	2.14E−04
× Dist. to nearest tier2			−9.92E−05	1.13E−04			−8.31e−04***	2.48E−04
× Dist. to nearest tier3			−3.93e−04***	1.22E−04			−1.71E−04	2.69E−04
× Dist. to nearest tier4			1.80E−04	2.11E−04			6.45E−04	4.60E−04
× Consumption facility density			−1.34e−05**	6.02E−06			5.69E−06	1.32E−05
× Central city			4.88e−05**	2.30E−05			1.39e−04***	5.00E−05
× Major road density			3.14e−04**	1.57E−04			−3.42E−04	3.48E−04
× Minor road density			1.26e−03***	2.25E−04			−1.09e−03**	5.02E−04
× Footway density			−6.65e−04***	2.58E−04			8.76E−04	5.81E−04
× Dist. to nearest subway			2.71E−04	2.43E−04			−1.09e−03**	5.33E−04
× Dist. to nearest bus			−5.42e−03*	3.04E−03			9.97E−03	6.64E−03
× No. of subway lines			1.36E−04	6.19E−04			−2.19e−03*	1.35E−03
× No. of bus lines			1.89E−06	6.59E−05			1.51E−04	1.45E−04
Constant	9.00e−01***	1.66E−01	7.92e−01**	3.22E−01	2.86e+00***	3.81E−01	1.64e+00**	7.13E−01
Lambda	7.51e−02***	3.86E−03	7.10e−02***	3.93E−03	1.17e−01***	2.66E−03	1.16e−01***	2.69E−03
Pseudo R-squared	0.212		0.231		0.184		0.201	

Note: # means the variable is log transformed.
***, ** and * represent the significance at the level of 99%, 95% and 90%, respectively.

is a nuanced relationship between distance to commercial centers and shopping travel distance. City-level and local community-level commercial center distances are positively correlated with shopping distance. This can be explained by the predominant distribution of city-level commercial centers in city cores. A greater distance to such centers might suggest a neighborhood's peripheral location, where residents typically undertake longer shopping trips (Jiao et al., 2016). On the other hand, local commercial centers predominantly cater to daily consumption needs. Thus, a greater distance to such centers necessitates longer travel for residents. Conversely, an increased distance to district and sub-district-level commercial centers relates to shorter shopping distances. This might be because residents, when distanced from these intermediate commercial hubs, opt for proximate local consumption outlets. Moreover, neighborhoods characterized by denser minor roads and those situated further from subway stations generally show longer shopping travel distances. In contrast, areas with dense consumption facilities and a higher number of bus lines tend to reduce the shopping travel distance for their residents.

Regarding the shopping trip frequency, we observe that closer proximity to commercial centers and higher densities of consumption facilities and major roads in neighborhoods encourage more frequent shopping trips. This finding is consistent with prior research (Cheng et al., 2019; Farag et al., 2007). Interestingly, areas with high land use diversity are associated with fewer shopping trips. A possible explanation for this counterintuitive outcome is that our dataset only records stays exceeding 30 min as an activity, potentially omitting brief daily consumption activities in highly diverse neighborhoods. This outcome could be further validated with a more granular dataset.

Additionally, the socio-demographics of neighborhoods do not appear to markedly influence travel behavior. This might be attributed to the aggregated nature of this study's focus: each neighborhood encapsulates a variety of socio-demographic profiles. We will delve deeper into the behavioral nuances across different socio-demographic subgroups in Section 4.4.

4.3. Moderating effects of E-shopping

To assess whether e-shopping mitigates the impact of land use factors on shopping trips, we introduce interaction terms between e-shopping usage and land use features. The results of the full model are

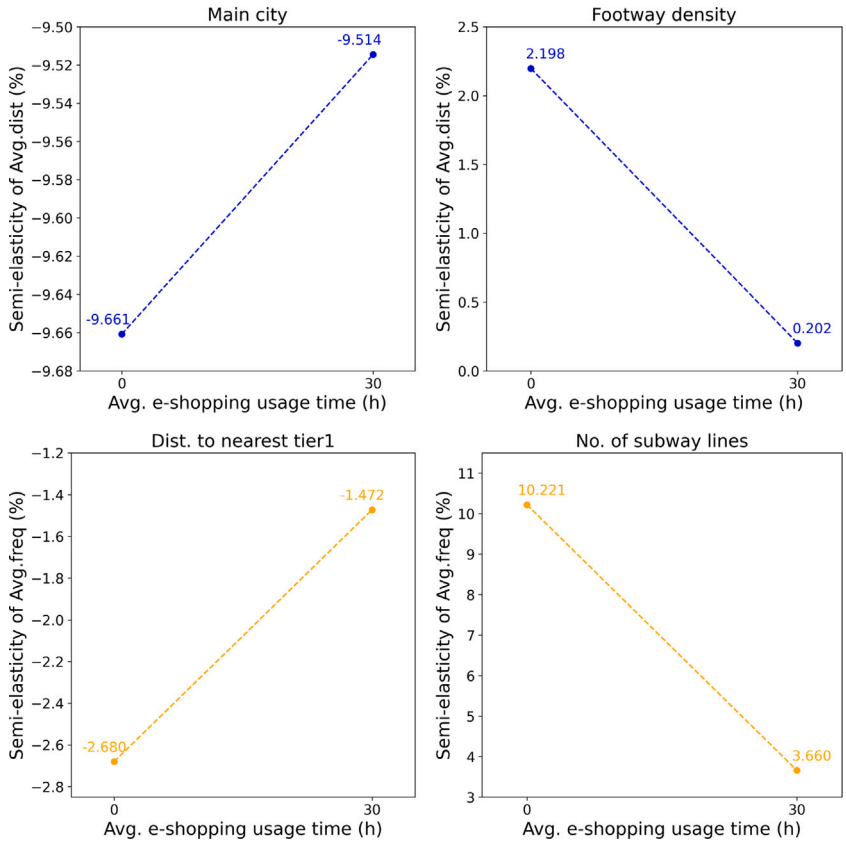


Fig. 5. Interaction of online shopping for traveling.

presented in Table 6. In this model, the effects of land use encompass both primary and interaction effects. Certain land use variables exhibit both significant primary and interaction effects, suggesting that e-shopping could potentially moderate their influence on travel patterns. In Table 6, we have distinctly marked these variables in blue for the distance model and in yellow for the frequency model.

Of these highlighted variables, most exhibit primary and interaction effect coefficients in opposite directions, with the sole exception being the distance to tier 1 commercial centers. This pattern underscores the moderating role e-shopping plays between land use attributes and travel behavior—primarily by diminishing the effect of land use features on offline shopping in both distance and frequency dimensions.

Specifically, in the absence of e-shopping, residing in the central city, higher densities of major and minor roads correlate with shorter shopping distances, whereas increased footway density correlates with longer distances. However, with the prevalence of e-shopping, these effects appear to be attenuated. Regarding shopping trip frequency, e-shopping might attenuate the adverse effects of greater distances to city-level commercial centers, while tempering the positive influences of proximity to district-level commercial centers and robust subway connectivity.

To better elucidate the mitigating effect of e-shopping, we quantify the specific effect values. Given that our dependent variables are log-transformed, it is straightforward to compute the semi-elasticity of dependent variables relative to influential factors. This measures the percentage change in the outcome in response to an absolute change in one independent variable. We graphically represent the semi-elasticity of shopping distance/frequency concerning land use features in Fig. 5. For clarity and aesthetically pleasing, only certain features are showcased.

In this figure, we depict the shifts in semi-elasticity as monthly e-shopping usage time varies from 0 to 30 h, approximating the average usage time. In the absence of e-shopping, residents of central urban regions typically have a shopping travel distance that is 9.66% shorter compared to their counterparts in non-central urban areas. However, when the e-shopping usage time reaches 30 h, this differential narrows to 9.51%. When considering footway density, an increase of 1 km/km^2 results in an average 2.20% surge in shopping distance without e-shopping. This increase is attenuated to 0.20% when e-shopping usage extends to 30 h. A similar trend is observed in shopping frequency: the semi-elasticity with respect to the proximity to city-level commercial centers shifts from −2.68% to −1.47%. Lastly, introducing an additional subway line correlates with a 10.22% boost in shopping frequency. This increment diminishes progressively with augmented online time, declining to 3.66% at a 30-h e-shopping usage rate.

Overall, e-shopping has a significant mitigating effect. This supports our hypothesis and generally indicates that the prevalence of online shopping is likely to attenuate the effectiveness of land use interventions. This result further corroborates our findings regarding the substitution effect of online shopping: as reliance on e-shopping increases, offline shopping decreases, consequently moderating the effects of land use characteristics on travel behavior.

4.4. Heterogeneity of moderating effects

Recognizing the unique demographic composition of different neighborhoods, it is critical to acknowledge that these characteristics might lead to diverse responses to online shopping (Mateos-Mínguez et al., 2021). Previous studies have found distinct shopping travel behavior across different age groups. For instance, elderly individuals tend to favor in-store shopping more than their younger counterparts (Bönisch et al., 2021). Daily shopping trips of older adults are also more restricted in terms of the spatial distribution of related facilities and individual mobility (Yang, Tang, Yang, Meng et al., 2022; Zhou et al., 2022).

Table 7
SEM results of senior-majority and senior-minority neighborhoods.

Variables	Avg. Dist #		Avg. Freq #	
	Senior-majority	Senior-minority	Senior-majority	Senior-minority
	Coef.	Coef.	Coef.	Coef.
ICT-related variables				
Avg. online time	1.48E−02	−8.34E−03	6.38E−04	5.35e−02*
Socio-demographics				
% of male	−9.01E−02	−3.84E−02	−2.34E−01	−1.79E−02
% <18 yrs old	1.32E−01	8.52E−02	9.15E−02	1.28E−01
% 18–39 yrs old	−2.77E−02	−6.82E−02	1.02e+00***	1.74E−01
% >60 yrs old	−2.45E−01	−9.96e−01*	1.29e+00**	9.89E−01
Avg. house price	−5.31E−04	1.36E−03	−1.55E−03	9.76e−03***
Land use features				
Residential density	7.15E−04	1.39E−03	8.44E−04	−4.00E−03
Job density	−2.56E−03	−5.55E−04	6.98e−03*	6.51E−03
Land use diversity	6.06e+00**	1.42E+00	−4.50E+00	4.36E+00
Dist. to nearest tier1	1.75e−02***	1.34e−02**	−2.54e−02*	−2.06e−02*
Dist. to nearest tier2	−5.60E−03	−2.49e−02***	7.16e−02***	1.39E−02
Dist. to nearest tier3	−7.12E−03	1.73e−02**	−3.98e−02*	5.65E−03
Dist. to nearest tier4	1.43E−02	−5.56E−03	−6.51e−02*	1.18E−02
Consuming facility density	−1.37E−04	2.36E−04	−1.10E−03	5.09E−04
Central city	−1.70e−01*	1.03E−01	−1.33E−01	−3.96E−01
Major road density	−2.17e−02**	−2.47e−02*	−1.47E−02	4.44E−02
Minor road density	−5.28e−02***	−5.35e−02***	6.08e−02*	6.35E−02
Footway density	6.22e−02***	2.64E−02	−2.30E−02	−4.60E−02
Dist. to nearest subway	−3.14e−02*	2.59e−02*	3.99E−02	−5.23E−03
Dist. to nearest bus	8.38E−02	2.21E−01	−1.06e+00*	−1.97E−01
No. of subway lines	−1.34E−02	9.15e−02*	1.64e−01*	−4.71E−02
No. of bus lines	−2.22E−03	−1.42e−02**	7.38E−03	7.09E−03
Interaction with Avg. online time				
× Residential density	−3.99E−05	−1.59E−05	1.87E−05	7.47E−05
× Job density	9.67e−05*	1.07E−05	−3.01e−04**	−1.64E−04
× Land use diversity	−9.37e−02*	−1.86E−02	−5.47E−02	−2.33e−01**
× Dist. to nearest tier1	1.07E−04	2.77E−04	4.26E−04	6.66e−04*
× Dist. to nearest tier2	−3.03E−04	2.22E−04	−1.95e−03***	−1.04e−03**
× Dist. to nearest tier3	1.97E−05	−6.38e−04***	7.81E−04	−1.64E−04
× Dist. to nearest tier4	4.00E−05	6.51e−04*	2.26e−03**	2.72E−04
× Consuming facility density	−2.68e−05**	−2.56e−05*	7.75e−05**	−4.59E−06
× Central city	7.21E−05	−1.05E−05	1.72E−04	1.93E−04
× Major road density	2.27E−04	5.62e−04*	5.28E−04	−1.09E−03
× Minor road density	1.80e−03***	2.17e−03***	−1.16E−03	−1.65E−03
× Footway density	−1.20e−03**	−9.92E−04	3.60E−04	1.36E−03
× Dist. to nearest subway	1.25e−03**	−1.88E−04	−2.69e−03**	−7.71E−04
× Dist. to nearest bus	−1.80E−03	−6.57E−03	3.52e−02**	1.59E−02
× No. of subway lines	1.09E−03	−2.28E−03	−4.88e−03*	6.39E−04
× No. of bus lines	1.01E−04	2.60e−04*	−1.15E−05	−1.34E−04
Constant	2.55E−01	1.49e+00***	2.09E+00	−1.96E−01
Lambda	6.81e−02***	9.61e−02***	1.29e−01***	1.50e−01***
Pseudo R-squared	0.278	0.227	0.261	0.208

A question then arises: does e-shopping exert varying moderating impacts on neighborhoods characterized by the proportion of elderly residents? To delve into this question, we categorize neighborhoods into two groups: senior-majority (those where the proportion of individuals aged 60 and above exceeds the 75th percentile) and senior-minority (those where this proportion falls below the 25th percentile). We apply SEMs on these two groups, and the results are compared and summarized in Table 7, where significant land use features and interaction terms are highlighted.

Our analysis reveals notable variability between these neighborhood types regarding their interactions with their immediate environment. First, the way in which land-use features shape travel behavior differs across these areas. Footway density, proximity to public transit, and local commercial centers are more influential for senior-majority neighborhoods, while the presence of extensive major/minor roads has a stronger impact on senior-minority neighborhoods.

These findings may be underpinned by age-related differences in travel mode preferences and perceptions of the built environment. Elderly individuals tend to rely more on walking or public transit travel for shopping, while younger adults prefer motorized modes such as cars (Zhou et al., 2022). For older individuals, who typically have restricted mobility, factors like footway density (where a higher density of pedestrian infrastructure correlates with better walkability (Lee et al., 2022; Yang, Tang, Yang, 2022)) become crucial in enabling longer travels. On the other hand, younger and middle-aged individuals might prioritize the road network's conditions, especially when driving. If their neighborhoods are saturated with major/minor roads, it might indicate limited parking availability and potential traffic congestion. Consequently, they might choose to walk or bike to nearby stores rather than drive to farther ones, resulting in shorter travel distances. Hence, urban practitioners and policy makers committed to promoting active travel are advised to pay more attention to those middle-aged and young adult gathered neighborhoods.

Second, our analysis reveals distinct moderating effects of e-shopping on the land use across these two types of neighborhoods. Neighborhoods predominantly occupied by seniors show a higher sensitivity to the mediation effect on travel frequency than those with a younger population, as indicated by a broader range of mitigated features. This can be attributed to two factors. First, elderly individuals may face more constraints in their daily shopping travels (Cheng et al., 2019), resulting in a more pronounced influence of the built environment on travel frequency in senior-majority areas. Second, elderly residents are less inclined to engage in multi-activity trip chains compared to younger adults (Golob & Hensher, 2007). Single-purpose shopping trips, prevalent among seniors, are more vulnerable

Table 8
The coefficients of Avg. online time in base models under all scenarios.

Models	Perturbation levels				
	0	5%	10%	15%	20%
Scenario 1 - Dependent variable uncertainty					
Avg. Dist model	−1.72E−03***	−1.83E−03***	−1.96E−03***	−2.11E−03***	−2.28E−03***
Avg. Freq model	−1.02E−02***	−1.03E−02***	−1.04E−02***	−1.06E−02***	−1.08E−02***
Scenario 2 - Avg. online time uncertainty					
Avg. Dist model	−1.72E−03***	−1.66E−03***	−1.55E−03***	−1.40E−03***	−1.24E−03***
Avg. Freq model	−1.02E−02***	−1.01E−02***	−9.72E−03***	−9.01E−03***	−8.15E−03***
Scenario 3 - Pandemic-related uncertainty					
Avg. Dist model	−1.72E−03***	−1.63E−03***	−1.51E−03***	−1.35E−03***	−1.19E−03***
Avg. Freq model	−1.02E−02***	−1.01E−02***	−9.69E−03***	−8.98E−03***	−8.13E−03***

to substitution by e-shopping compared to multi-activity trip chains. As a result, senior-majority neighborhoods might experience a wider range of mitigating effects from various land use features. In contrast, senior-minority neighborhoods might display more stability in their shopping behavior, being less influenced by e-shopping.

In conclusion, the differences in travel mode preferences and perceptions of the built environment lead to distinct shopping distances and frequencies between these two neighborhood types. Additionally, the reduced inclination of seniors to form multi-activity trip chains heightens the susceptibility of senior-majority neighborhoods to the mitigating effects of e-shopping.

4.5. Robustness analysis

In our robustness experiments, we evaluate the stability of the model results under three scenarios as introduced in Table 3. In each scenario, we sample the perturbation term ϵ from a normal distribution with standard deviations of 5%, 10%, 15%, and 20%, respectively.

To examine the robustness of the substitution effect, Table 8 reports the coefficients of *Avg. online time* in the baseline models across all scenarios. The results show that Average online usage time consistently exhibits a negative and statistically significant coefficient in all scenarios, although the magnitude of the coefficients varies slightly. This implies that individual-level data biases do not affect the overall conclusion regarding the substitution effect at the community level.

Figs. 6 to 11 further illustrate the coefficients and significance levels of the full models and subgroup models under different perturbation levels. In each figure, the first row presents coefficient estimates while the second row shows the corresponding significance levels. The first column displays results from the full models of *Avg. Dist*/*Avg. Freq*, while the second and third columns present the subgroup model results. For clarity, we only plot variables highlighted in Tables 6 and 7.

For the full models, the coefficients remain stable across different perturbation levels: no coefficient changes sign and most remain statistically significant. Only two variables show noticeable changes. First, the interaction term between Average online time and Main city in the *Avg. Dist* model under Scenario 1 becomes less significant when the standard deviation of the error rate reaches 0.2 (see Fig. 6). Second, the interaction term between Average online time and Number of subway lines in the *Avg. Freq* model under Scenario 1 becomes marginally insignificant at the 10% level (see Fig. 7). These results suggest that our conclusions regarding the attenuating effects remain largely robust under different perturbation levels.

For the subgroup models, the signs of coefficients remain stable across all perturbation levels, although the significance levels exhibit some fluctuation, especially in the senior-minority models. For example, as shown in Fig. 6, when the standard deviation of the error rate exceeds 5% under Scenario 1, the impact of Major road density and its interaction term on shopping trip distance in senior-minority communities becomes insignificant. A similar trend is observed under Scenario 3 (see Fig. 10). This indicates that, under shopping trip misclassification and pandemic-related uncertainties, the moderating effect of e-shopping on the influence of major road density in senior-minority areas becomes less reliable, and its interpretation should be more cautious and conservative.

Overall, our robustness analysis demonstrates that uncertainty in the translation from Average online time to actual purchases has minimal impact on the conclusions. While misclassification of shopping trips and pandemic effects influence certain variables, most findings remain robust. Variables that exhibit greater sensitivity to perturbations should be further examined in future research when more suitable data become available.

5. Conclusion

Utilizing a one-month sample of mobile signaling data, this study probes the question of whether online shopping activities attenuate the effects of neighborhood land use on shopping travel distance and frequency. Using Spatial Error Models, we find that attenuating effects are evident at the neighborhood level. This shows that a reliance on e-shopping can reduce the influence of land use features such as proximity to commercial centers, geographical location, road density, and public transit accessibility on travel behavior. Notably, these effects vary across neighborhood types, with senior-majority neighborhoods demonstrating greater sensitivity to attenuating effects on travel frequency than senior-minority neighborhoods. Furthermore, senior-majority neighborhoods are more influenced by factors such as footway density, proximity to public transit, and local commercial centers, whereas senior-minority neighborhoods are affected by the density of major/minor roads. These distinctions may be attributed to differences in travel mode preferences, built environment perceptions, and trip chain generation between older and younger adults. Additionally, our results show that stronger e-shopping engagement decreases the number of shopping trips (i.e., exhibits a substitution effect) and modifies travel distance.

This research contributes to the existing literature in three respects. First, it affirms the moderating role of e-shopping at the neighborhood level, particularly in attenuating the influence of land use factors on shopping trip distance and frequency. This aspect has not been thoroughly explored in previous research within the realm of shopping travel behavior analysis. Second, the investigation of heterogeneity in attenuating effects across neighborhoods of varying demographic compositions reveals more intricate effects within different sample subsets. Our study underscores the importance of considering population segmentation when examining the adjustment to online shopping and contributes to the literature on shopper heterogeneity, which has historically received limited attention (Le et al., 2022). Last, our study unveils the coexistence of both substitution and modification relationships between online and offline shopping, using mobile signaling data for the first time for this research topic. Earlier research predominantly relied on surveys, which offer valuable behavioral insights but are often constrained in spatial coverage and observation periods.

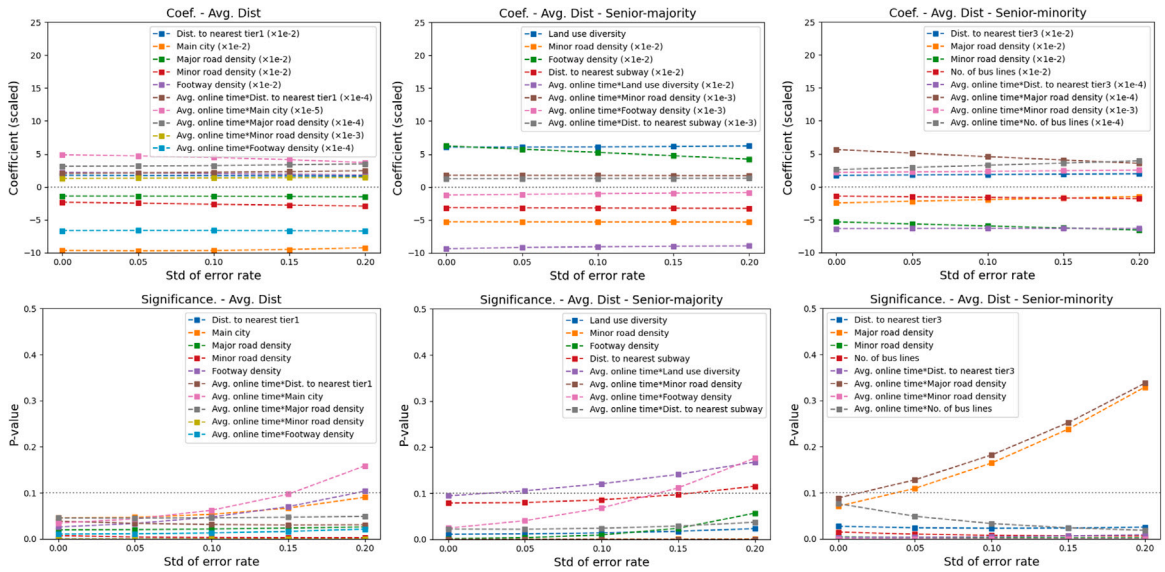


Fig. 6. The Avg. Dist model outcomes of coefficients and significance at various perturbation levels under scenario 1.

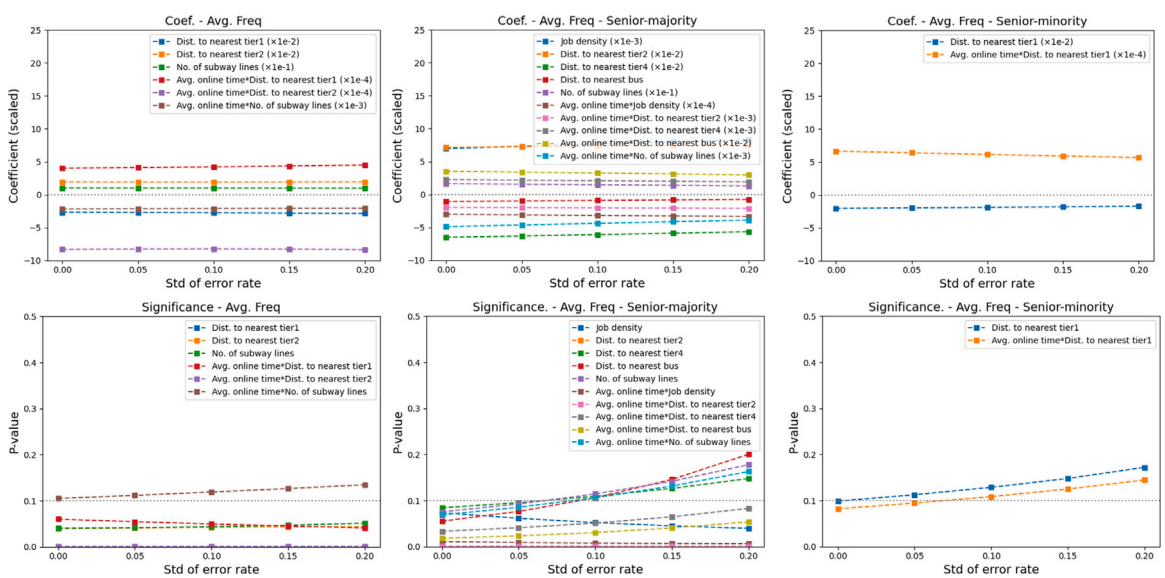


Fig. 7. The Avg. Freq model outcomes of coefficients and significance at various perturbation levels under scenario 1.

In contrast, our dataset, containing detailed e-shopping app usage, land use features, neighborhood demographics, and travel information, serves as a valuable complement to existing survey-based research. Rather than replacing traditional methods, our approach provides a new perspective and analytical possibility for understanding evolving shopping behaviors in the digital era.

6. Discussion

6.1. The implications of conclusions

The insights of this work can draw important implications for urban transportation system management and community planning. Our study generally supports the hypothesis of a weakened neighborhood land use-travel relationship due to e-shopping, which poses a challenge for travel demand management. Traditional land use interventions might be less effective amidst the prevalence of online shopping activities. Urban planners should therefore account for e-shopping impacts when modeling travel demand and formulating land use policies. Additionally, it is highlighted that implementing land use interventions requires consideration of community demographic heterogeneity. For senior-majority communities, enhancing walkability and public transportation accessibility around the community is crucial for meeting the basic travel needs of the elderly (Zhou et al., 2022). For communities primarily inhabited by middle-aged and younger people, improving non-motorized road conditions and ensuring consumption facility accessibility around the community is essential. This will ensure residents' travel capabilities, convert residents' long-distance travel needs (especially non-active travel) into short-range consumption trips, and potentially alter residents' travel modes.

From an urbanization perspective, the widespread adoption of e-shopping is likely to reshape the configuration of cities and the distribution of urban activities (Colaço & e Silva, 2025). As land use factors become less influential, consumption facilities and transport accessibility — both of which can be substituted by online channels — may exert a weaker role in residential location choices. Consequently, households may prefer larger dwellings in suburban or peripheral areas, thereby reinforcing metropolitan decentralization processes. In

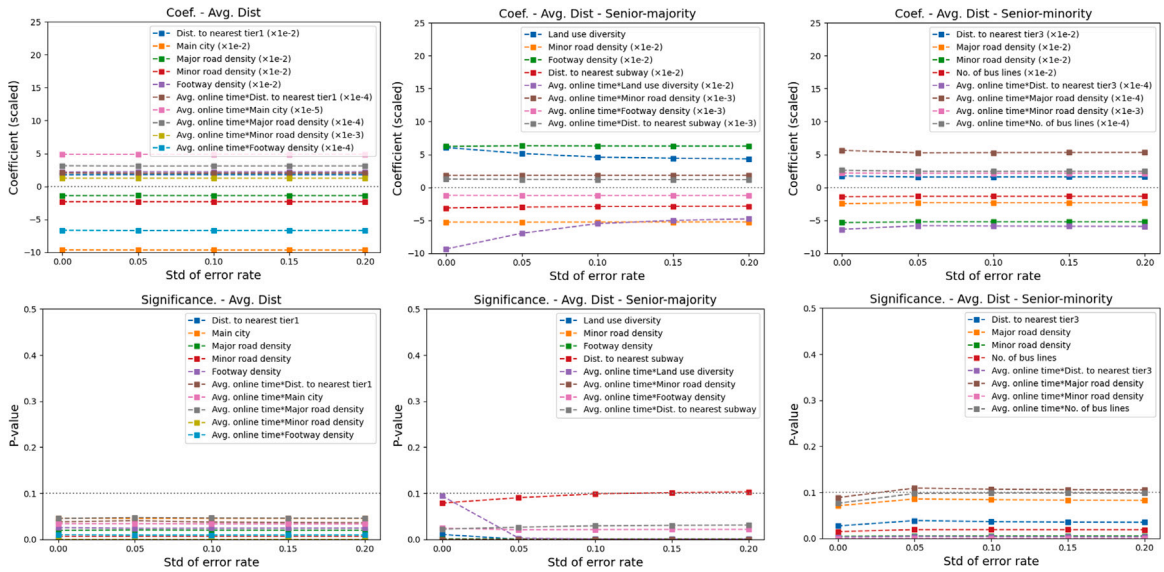


Fig. 8. The Avg. Dist model outcomes of coefficients and significance at various perturbation levels under scenario 2.

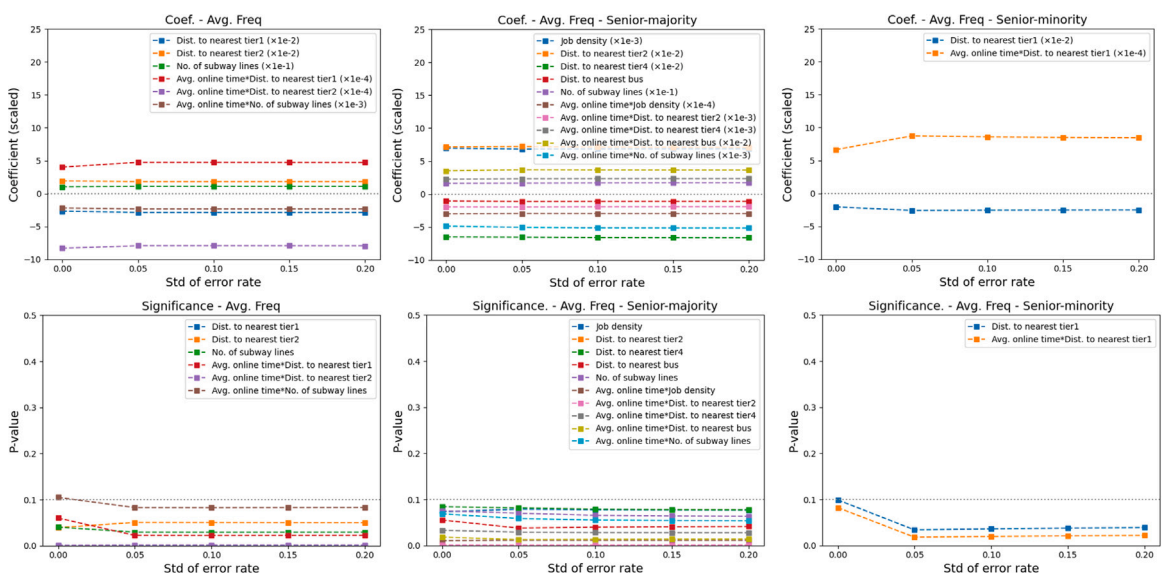


Fig. 9. The Avg. Freq model outcomes of coefficients and significance at various perturbation levels under scenario 2.

addition, a center–periphery relocation of retail activity may become more pronounced as a response to rising commercial rents and the substitution effects of e-shopping (Colaço & e Silva, 2025). Such dynamics could undermine policies of the past two decades that aimed to promote compact urban development and contain urban sprawl (Arranz-López & Soria-Lara, 2022).

Moreover, this study suggests that e-shopping might mitigate travel demand due to its substitution effect on shopping trips. Traditionally, policy-makers have focused on improving the level of service of public transportation systems to encourage a modal shift away from private car use (Bibri et al., 2020; Moreno et al., 2021). In the digital era, however, e-shopping offers an innovative opportunity to alleviate traffic burden by providing convenient delivery services for daily shopping needs. Furthermore, residents living in areas with limited access to public transportation can still obtain essential goods through e-shopping, thereby enhancing their accessibility without requiring frequent travel. This highlights the potential for greater flexibility in urban transportation planning, particularly when addressing challenging scenarios such as land use constraints or vulnerabilities to extreme weather conditions.

Nevertheless, while e-shopping has the potential to mitigate passenger travel demand, it may simultaneously generate additional freight trips and greater freight transport distances (Anderson et al., 2003; Rotem-Mindali & Salomon, 2007; Weltevreden & Rotem-Mindali, 2009). These changes can have significant consequences for residential areas, including noise, pedestrian safety risks, and curbside occupation (Colaço & e Silva, 2025). In addition, the net impact of e-shopping on traffic congestion should consider the dominant mode of shopping travel. If active travel methods such as walking and biking predominantly constitute these trips, then it is unlikely to result in a reduction of personal vehicle use or alleviate traffic congestion. Instead, it could lead to further transportation challenges due to increased delivery trips (Xi et al., 2020). Further exploration of this issue would be possible if detailed data on shopping travel modes were available. Furthermore, issues arising from the dependence on online activities, such as local commerce detriment, residential neighborhood vitality reduction, and increased social isolation, also merit attention.

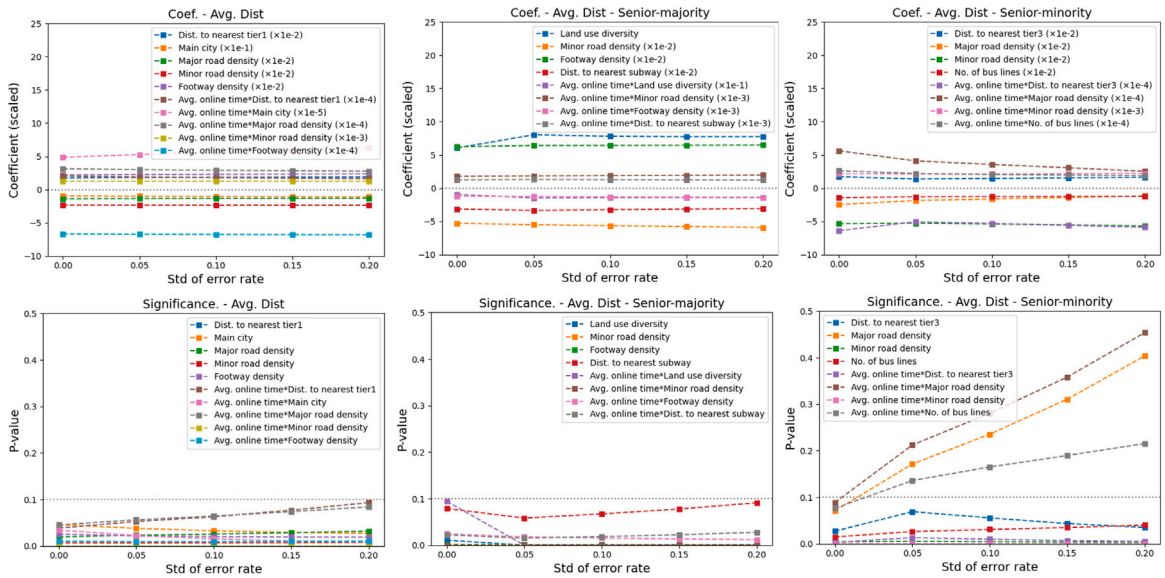


Fig. 10. The Avg. Dist model outcomes of coefficients and significance at various perturbation levels under scenario 3.

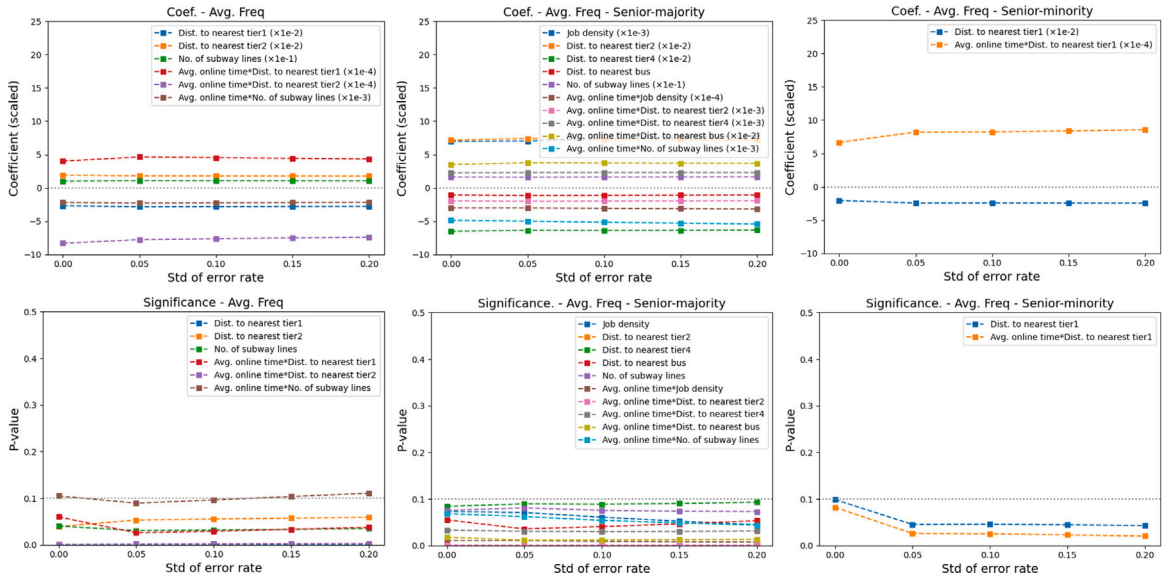


Fig. 11. The Avg. Freq model outcomes of coefficients and significance at various perturbation levels under scenario 3.

6.2. The limitation of mobile phone data

Although mobile phone data can complement traditional survey data in many aspects, it should be acknowledged that such data also present several limitations. We summarize these limitations and potential solutions in Table 9.

With respect to common issues, the first limitation is that mobile phone data lack specific trip attributes, such as travel purpose. Algorithms are therefore required to infer these attributes. In this study, we applied a rule-based POI matching method to identify shopping trips and validated the accuracy of this method using travel survey data. More advanced approaches, such as combining probabilistic matching and users' travel history patterns, or machine learning algorithm (Alsger et al., 2018; Gong et al., 2018; Xiao et al., 2016), may achieve higher accuracy; however, no algorithm can guarantee perfect results. To address this, we incorporated robustness analysis to account for data uncertainty and to test the stability of our model results.

Second, according to documents provided by the network operator, some users may purchase multiple SIM cards for family use, which can bias the gender and age information associated with individual cards—particularly for elderly and female users. To mitigate this, network operators typically calibrate user-level demographic data against local census statistics, and then assign demographic weights to users. Weighted aggregation thus allows us to obtain less biased population statistics at the neighborhood level.

Third, compared with survey data, mobile phone data contain fewer individual attributes. In our dataset, only age and gender are available, and these may still be biased as noted above. Therefore, in this study we used neighborhood-level population data as a proxy and adopted housing prices as an indicator of neighborhood income levels.

Fourth, mobile phone data do not include actual e-shopping information such as transaction amount or product type. Such data can only be obtained from app operators. Even if purchase data were accessible from app, the anonymity of mobile phone data makes it nearly impossible to match individual mobile users with app users. Consequently, we assumed that online usage time is positively associated with online purchasing, and we conducted robustness analysis to simulate uncertainty in the transfer rate. Another possible approach

Table 9
The limitation of mobile phone data and potential solutions.

Type	Limitation	Potential solution
Common issues	Unknown trip context (e.g., trip purpose)	Apply more accurate inference methods; validate results with auxiliary data; incorporate uncertainty quantification
	Underrepresentation of specific groups (e.g., older adults, women)	Use sample-weighted statistics to adjust for demographic bias
	Limited socio-demographic information	Employ neighborhood-level statistics as a proxy
	Unknown actual online purchasing behavior	Validate with additional auxiliary datasets; incorporate uncertainty quantification
Case-specific	Aggregated data	Conduct analysis at the neighborhood level
	Computational constraints in online platforms	Request higher-level access; purchase customized computing services; incorporate uncertainty quantification
	Pandemic-induced disruptions	Apply uncertainty quantification to account for abnormal patterns; use post-pandemic data

Note: Solutions shown with underlining indicate those applied in our study.

would be to collect survey data linking online usage time with actual purchasing behavior to better estimate transfer rates. Moreover, given that shopping apps today often serve multiple functions — such as food delivery in addition to retail shopping — it is not feasible to distinguish between dining and shopping based on app usage alone. Therefore, this study focuses on consumption in a broader sense.

In addition to these general issues, some limitations are case-specific to our study. First, due to privacy constraints, the mobile phone data were provided in aggregated form, preventing individual-level modeling. We therefore conducted analysis at the neighborhood level. Second, owing to the platform regulations, individual-level data processing could only be performed online within the secure platform environment. Once aggregated, the data could be downloaded for further modeling and analysis. However, because the platform is SQL-based, only simple rule-based processing could be implemented at the individual level, limiting the use of more advanced inference algorithms. One possible solution would be to purchase customized computing services from the network provider, but in this study we instead relied on robustness analysis. Third, our dataset was collected during the COVID-19 pandemic. Our robustness analysis indicates that pandemic-related disruptions did not significantly affect our main findings regarding the substitution effect of e-shopping and the attenuated impact of land use. Moreover, the behavioral changes observed during the pandemic may partly reflect patterns that persist into the post-pandemic era (Colaço & de Abreu e Silva, 2024). Nevertheless, future research should utilize more recent mobile phone data to further validate these conclusions.

Finally, although mobile phone data have certain limitations, this study represents a valuable complement and preliminary attempt to extend the existing e-shopping literature using such data. Mobile phone data cannot fully replace the advantages of survey data in this field. However, we hope that more comprehensive datasets will become available in the future to further test and strengthen our conclusions.

CRediT authorship contribution statement

Fangyi Ding: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Zhan Zhao:** Writing – review & editing, Methodology, Conceptualization. **Yun Han:** Data curation, Validation. **Yang Zhou:** Writing – review & editing. **Yang Xu:** Writing – review & editing.

Declaration of competing interest

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

Acknowledgments

The authors would like to thank China Unicom for providing the data that makes this study possible. This research is supported by the Seed Funding for Strategic Interdisciplinary Research Scheme (102010057) at the University of Hong Kong, Hong Kong.

Appendix A. The validation of shopping trip inference

To validate the accuracy of the POI–activity matching method, we use independent travel survey data as a benchmark. The travel survey was conducted in Shanghai in 2019 and includes travel records from 51,114 households and 111,511 individuals. Since each trip record contains detailed information on departure/arrival time, location, and travel purpose, it provides a reliable ground truth for validation.

Specifically, we extract home-based non-commuting trips from the survey data, using the same criteria as in the mobile signaling-based approach. For each trip destination, we assign the nearest POI and obtain an estimated POI-type label, denoted as $\hat{L}$, while the actual activity purpose from the survey serves as the ground truth label, denoted as L. To assess the accuracy of shopping trip inference, we calculate the proportion of trips labeled as shopping in the ground truth (L_s) among those identified as commercial POIs ($\hat{L}_s$). The accuracy is computed as:

$$Acc = \frac{L_s}{\hat{L}_s} \tag{A.1}$$

The validation results show an accuracy of 82.3%, demonstrating that this POI-matching method can achieve high reliability in inferring shopping activities.

Appendix B. The standard deviation of error rate at the neighborhood level

We define the recognition outcome of each trip in grid m as a Bernoulli random variable:

$$X_{im} \sim \text{Bernoulli}(q), \quad P(X_{im} = 1) = q, \quad P(X_{im} = 0) = 1 - q, \tag{B.1}$$

where q denotes the probability of correct classification.

The average recognition accuracy in grid m is:

$$\bar{X}_m = \frac{1}{n_m} \sum_{i=1}^{n_m} X_{im}. \tag{B.2}$$

By the Central Limit Theorem, when n_m is sufficiently large, $\bar{X}_m$ approximately follows a normal distribution with:

$$E[\bar{X}_m] = q, \qquad \text{Var}(\bar{X}_m) = \frac{pq}{n_m}, \quad p = 1 - q. \tag{B.3}$$

Thus, the standard deviation of $\bar{X}_m$ is:

$$\sigma_{\bar{X}_m} = \sqrt{\frac{pq}{n_m}}. \tag{B.4}$$

If we define the relative error as:

$$\tilde{\epsilon}_m = \frac{\bar{X}_m - q}{q}, \tag{B.5}$$

then the corresponding standard deviation is:

$$\sigma_{\tilde{\epsilon}_m} = \frac{\sqrt{pq/n_m}}{q}. \tag{B.6}$$

References

Aldo, A.-L., Andreas, B., & Martin, L. (2023). Exploring the associations between E-shopping and the share of shopping trip frequency and travelled time over total daily travel demand. *Travel Behaviour and Society*, *31*, 202–208.

Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., & Hickman, M. (2018). Public transport trip purpose inference using smart card fare data. *Transportation Research Part C: Emerging Technologies*, *87*, 123–137.

Anderson, W. P., Chatterjee, L., & Lakshmanan, T. (2003). E-commerce, transportation, and economic geography. *Growth and Change*, *34*(4), 415–432.

Anselin, L. (1988). *Spatial econometrics: Methods and models*: *vol. 4*, Springer Science & Business Media.

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*(1), 77–104.

Arranz-López, A., & Soria-Lara, J. A. (2022). ICT use and spatial fragmentation of activity participation in post-COVID-19 urban societies. *Land Use Policy*, *120*, Article 106302.

Beckers, J., Cárdenas, I., & Verhetsel, A. (2018). Identifying the geography of online shopping adoption in Belgium. *Journal of Retailing and Consumer Services*, *45*, 33–41.

Bergroth, C., Järv, O., Tenkanen, H., Manninen, M., & Toivonen, T. (2022). A 24-hour population distribution dataset based on mobile phone data from Helsinki Metropolitan Area, Finland. *Scientific Data*, *9*(1), 39.

Bibri, S. E., Krogstie, J., & Kärrholm, M. (2020). Compact city planning and development: Emerging practices and strategies for achieving the goals of sustainability. *Developments in the Built Environment*, *4*, Article 100021.

Boarnet, M., & Crane, R. (2001). The influence of land use on travel behavior: specification and estimation strategies. *Transportation Research Part A: Policy and Practice*, *35*(9), 823–845.

Boarnet, M. G., Joh, K., Siembab, W., Fulton, W., & Nguyen, M. T. (2011). Retrofitting the suburbs to increase walking: Evidence from a land-use-travel study. *Urban Studies*, *48*(1), 129–159.

Bönisch, L., von Behren, S., Chlond, B., & Vortisch, P. (2021). Insights into shopping travel behavior: latent classes in relation to attitudes towards shopping. *European Transport Research Review*, *13*(1), 1–13.

Calderwood, E., & Freathy, P. (2014). Consumer mobility in the Scottish isles: The impact of internet adoption upon retail travel patterns. *Transportation Research Part A: Policy and Practice*, *59*, 192–203.

Cao, X. (2009). E-shopping, spatial attributes, and personal travel: a review of empirical studies. *Transportation Research Record*, *2135*(1), 160–169.

Cao, X., Chen, Q., & Choo, S. (2013). Geographic distribution of e-shopping: application of structural equation models in the Twin Cities of Minnesota. *Transportation Research Record*, *2383*(1), 18–26.

Cao, X., Douma, F., & Cleaveland, F. (2010). Influence of E-shopping on shopping travel: Evidence from Minnesota's Twin Cities. *Transportation Research Record*, *2157*(1), 147–154.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219.

Cheng, L., Chen, X., Yang, S., Cao, Z., De Vos, J., & Witlox, F. (2019). Active travel for active ageing in China: The role of built environment. *Journal of Transport Geography*, *76*, 142–152.

Clarke, G., Thompson, C., & Birkin, M. (2015). The emerging geography of e-commerce in British retailing. *Regional Studies, Regional Science*, *2*(1), 371–391.

Colaco, R., & de Abreu e Silva, J. (2021). Exploring the interactions between online shopping, in-store shopping, and weekly travel behavior using a 7-day shopping survey in Lisbon, Portugal. *Transportation Research Record*, *2675*(5), 379–390.

Colaço, R., & e Silva, J. d. (2022). Exploring the e-shopping geography of lisbon: Assessing online shopping adoption for retail purchases and food deliveries using a 7-day shopping survey. *Journal of Retailing and Consumer Services*, *65*, Article 102859.

Colaço, R., & de Abreu e Silva, J. (2024). Interactions between online shopping, in-store shopping and weekly travel behavior: An analysis before and in the aftermath of COVID-19 in Lisbon, Portugal. *Transportation Research Record*, *2678*(12), 2112–2124.

Colaço, R., & e Silva, J. d. (2025). Are we moving online? Assessing the interactions between telework and grocery purchases, eating out and meal deliveries. *Transportation Research Part A: Policy and Practice*, *195*, Article 104461.

Congress for the New Urbanism (2024). The charter of the New Urbanism. URL: https://www.cnu.org/who-we-are/charter-new-urbanism.

Dhananjaya, D., & Sivakumar, T. (2022). Enhancing the POI data for trip purpose inference using machine learning techniques. In *2022 IEEE 25th international conference on intelligent transportation systems* (pp. 3496–3501). IEEE.

Ding, Y., & Lu, H. (2017). The interactions between online shopping and personal activity travel behavior: an analysis with a GPS-based activity travel diary. *Transportation*, *44*, 311–324.

Ding, F., Tang, Y., Wang, Y., & Zhao, Z. (2025). Unraveling the network effects in station ridership growth patterns under metro network expansion. *Journal of Transport Geography*, *125*, Article 104205.

Ding, C., Wang, D., Liu, C., Zhang, Y., & Yang, J. (2017). Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A: Policy and Practice*, *100*, 65–80.

Elldér, E. (2017). Does telework weaken urban structure–travel relationships? *Journal of Transport and Land Use*, *10*(1), 187–210.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *vol. 96*, In *Kdd* (pp. 226–231).

Etminani-Ghasrodashti, R., & Hamidi, S. (2020). Online shopping as a substitute or complement to in-store shopping trips in Iran? *Cities*, *103*, Article 102768.

Ewing, R., & Cervero, R. (2001). Travel and the built environment: a synthesis. *Transportation Research Record*, *1780*(1), 87–114.

Farag, S., Schwanen, T., & Dijst, M. (2005). Empirical investigation of online searching and buying and their relationship to shopping trips. *Transportation Research Record*, *1926*(1), 242–251.

Farag, S., Schwanen, T., Dijst, M., & Faber, J. (2007). Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. *Transportation Research Part A: Policy and Practice*, *41*(2), 125–141.

Farag, S., Weltevreden, J., Van Rietbergen, T., Dijst, M., & van Oort, F. (2006). E-shopping in the Netherlands: Does geography matter? *Environment and Planning B: Planning and Design*, *33*(1), 59–74.

Ferrell, C. E. (2004). Home-based teleshoppers and shopping travel: Do teleshoppers travel less? *Transportation Research Record*, *1894*(1), 241–248.

Ferrell, C. E. (2005). Home-based teleshopping and shopping travel: Where do people find the time? *Transportation Research Record*, *1926*(1), 212–223.

Gao, X., Liu, L., Ou, D., & Yuwu, H. (2023). Using mobile phone data to probe the mobility-related well-being of rural residents: A case study of Chengdu, China. *Regional Science Policy & Practice*, *15*(9), 1893–1907.

Golob, T. F., & Hensher, D. A. (2007). The trip chaining activity of Sydney residents: a cross-section assessment by age group with a focus on seniors. *Journal of Transport Geography*, *15*(4), 298–312.

Gong, L., Kanamori, R., & Yamamoto, T. (2018). Data selection in machine learning for identifying trip purposes and travel modes from longitudinal GPS data collection lasting for seasons. *Travel Behaviour and Society*, *11*, 131–140.

He, X., & He, S. Y. (2024). Using mobile phone big data and street view images to explore the mismatch between walkability and walking behavior. *Transportation Research Part A: Policy and Practice*, *180*, Article 103946.

Jiao, J., Vernez Moudon, A., & Drewnowski, A. (2016). Does urban form influence grocery shopping frequency? A study from Seattle, Washington, USA. *International Journal of Retail & Distribution Management*, *44*(9), 903–922.

Kwan, M.-P., Dijst, M., & Schwanen, T. (2007). The interaction between ICT and human activity-travel behavior. *Transportation Research. Part A, Policy and Practice*, *41*(2), 121–124.

Le, H. T., Carrel, A. L., & Shah, H. (2022). Impacts of online shopping on travel demand: a systematic review. *Transport Reviews*, *42*(3), 273–295.

Lee, R. J., Sener, I. N., Mokhtarian, P. L., & Handy, S. L. (2017). Relationships between the online and in-store shopping frequency of Davis, California residents. *Transportation Research Part A: Policy and Practice*, *100*, 40–52.

Lee, J., Vojnovic, I., & Grady, S. C. (2022). What lies beyond the built environment? Exploring "walkability" across the Detroit region. *Applied Geography*, *141*, Article 102670.

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Li, J., & Kim, C. (2020). Exploring relationships of grocery shopping patterns and healthy food accessibility in residential neighborhoods and activity space. *Applied Geography*, *116*, Article 102169.

Loo, B. P., & Wang, B. (2018). Factors associated with home-based e-working and e-shopping in Nanjing, China. *Transportation*, *45*(2), 365–384.

Maat, K., & Konings, R. (2018). Accessibility or innovation? Store shopping trips versus online shopping. *Transportation Research Record*, *2672*(50), 1–10.

Mateos-Mínguez, P., Arranz-López, A., Soria-Lara, J. A., & Lanzendorf, M. (2021). E-shoppers and multimodal accessibility to in-store retail: An analysis of spatial and social effects. *Journal of Transport Geography*, *96*, Article 103198.

Meng, C., Cui, Y., He, Q., Su, L., & Gao, J. (2017). Travel purpose inference with GPS trajectories, POIs, and geo-tagged social media data. In *2017 IEEE international conference on big data* (pp. 1319–1324). IEEE.

Moreno, C., Allam, Z., Chabaud, D., Gall, C., & Pratlong, F. (2021). Introducing the "15-Minute City": Sustainability, resilience and place identity in future post-pandemic cities. *Smart Cities*, *4*(1), 93–111.

Oosterlinck, D., Benoit, D. F., Baecke, P., & Van de Weghe, N. (2017). Bluetooth tracking of humans in an indoor environment: An application to shopping mall visits. *Applied Geography*, *78*, 55–65.

Ramezani, S., Laatikainen, T., Hasanzadeh, K., & Kyttä, M. (2021). Shopping trip mode choice of older adults: an application of activity space and hybrid choice models in understanding the effects of built environment and personal goals. *Transportation*, *48*(2), 505–536.

Ren, F., & Kwan, M.-P. (2009). The impact of geographic context on e-shopping behavior. *Environment and Planning B: Planning and Design*, *36*(2), 262–278.

Rotem-Mindali, O., & Salomon, I. (2007). The impacts of E-retail on the choice of shopping trips and delivery: Some preliminary findings. *Transportation Research Part A: Policy and Practice*, *41*(2), 176–189.

Rotem-Mindali, O., & Weltevreden, J. W. (2013). Transport effects of e-commerce: what can be learned after years of research? *Transportation*, *40*, 867–885.

Shanghai Municipal People's Government (2018). Shanghai master plan 2017–2035.

Shao, R., Derudder, B., Yang, Y., & Witlox, F. (2023). The association between transit accessibility and space-time flexibility of shopping travel: On the moderating role of ICT use. *Journal of Transport Geography*, *111*, Article 103661.

Shi, K., Cheng, L., De Vos, J., Yang, Y., Cao, W., & Witlox, F. (2021). How does purchasing intangible services online influence the travel to consume these services? A focus on a Chinese context. *Transportation*, *48*(5), 2605–2625.

Shi, K., De Vos, J., Yang, Y., Li, E., & Witlox, F. (2020). Does e-shopping for intangible services attenuate the effect of spatial attributes on travel distance and duration? *Transportation Research Part A: Policy and Practice*, *141*, 86–97.

Shi, K., De Vos, J., Yang, Y., & Witlox, F. (2019). Does e-shopping replace shopping trips? Empirical evidence from Chengdu, China. *Transportation Research Part A: Policy and Practice*, *122*, 21–33.

Shi, K., Shao, R., De Vos, J., Cheng, L., & Witlox, F. (2021). Is e-shopping likely to reduce shopping trips for car owners? A propensity score matching analysis. *Journal of Transport Geography*, *95*, Article 103132.

Shi, K., Shao, R., De Vos, J., & Witlox, F. (2023). Do e-shopping attitudes mediate the effect of the built environment on online shopping frequency of e-shoppers? *International Journal of Sustainable Transportation*, *17*(1), 41–51.

Sim, L. L., & Koi, S. M. (2002). Singapore's Internet shoppers and their impact on traditional shopping patterns. *Journal of Retailing and Consumer Services*, *9*(2), 115–124.

Su, F., Schmöcker, J.-D., & Bell, M. G. (2009). Mode choice of older people before and after shopping: A study with London data. *Journal of Transport and Land Use*, *2*(1), 29–46.

Vij, A., & Shankari, K. (2015). When is big data big enough? Implications of using GPS-based surveys for travel demand analysis. *Transportation Research Part C: Emerging Technologies*, *56*, 446–462.

Wei, Z., Tang, Q., & Zhen, F. (2024). Spatial and social heterogeneities of residents' online shopping behaviors within a large Chinese city: The case of Weifang. *Applied Geography*, *167*, Article 103289.

Weltevreden, J. W., & Rotem-Mindali, O. (2009). Mobility effects of b2c and c2c e-commerce in the Netherlands: a quantitative assessment. *Journal of Transport Geography*, *17*(2), 83–92.

Xi, G., Cao, X., & Zhen, F. (2020). The impacts of same day delivery online shopping on local store shopping in Nanjing, China. *Transportation Research Part A: Policy and Practice*, *136*, 35–47.

Xi, G., Zhen, F., Gilles, P., & Valerie, F. (2017). Spatio-temporal fragmentation of leisure activities in information era: Empirical evidence from Nanjing, China. *Chinese Geographical Science*, *27*, 137–150.

Xiao, G., Juan, Z., & Zhang, C. (2016). Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization. *Transportation Research Part C: Emerging Technologies*, *71*, 447–463.

Xie, K., Deng, K., & Zhou, X. (2009). From trajectories to activities: a spatio-temporal join approach. In *Proceedings of the 2009 international workshop on location based social networks* (pp. 25–32).

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z., & Li, Q. (2015). Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. *Transportation*, *42*(4), 625–646.

Yang, C., Tang, X., & Yang, L. (2022). Spatially varying associations between the built environment and older adults' propensity to walk. *Frontiers in Public Health*, *10*, Article 1003791.

Yang, L., Tang, X., Yang, H., Meng, F., & Liu, J. (2022). Using a system of equations to assess the determinants of the walking behavior of older adults. *Transactions in GIS*, *26*(3), 1339–1354.

Zahnow, R., Kimpton, A., Corcoran, J., & Mielke, G. (2022). Neighbourhood correlates of average population walking: using aggregated, anonymised mobile phone data to identify where people walk. *Health & Place*, *77*, Article 102892.

Zhen, F., Du, X., Cao, J., & Mokhtarian, P. L. (2018). The association between spatial attributes and e-shopping in the shopping process for search goods and experience goods: Evidence from Nanjing. *Journal of Transport Geography*, *66*, 291–299.

Zhou, Y., & Wang, X. C. (2014). Explore the relationship between online shopping and shopping trips: An analysis with the 2009 NHTS data. *Transportation Research Part A: Policy and Practice*, *70*, 1–9.

Zhou, Y., Yuan, Q., Ding, F., Chen, M., Yang, C., & Guo, T. (2022). Demand, mobility, and constraints: Exploring travel behaviors and mode choices of older adults using a facility-based framework. *Journal of Transport Geography*, *102*, Article 103368.

Zhou, Y., Yuan, Q., Yang, C., Guo, T., Ma, X., Sun, W., & Yang, T. (2024). Residence-workplace identification and validation based on mobile phone data: A case study in a large-scale urban agglomeration in China. *Transportation Research Record*, Article 03611981241270163.

Zhu, X., & Wu, X. (2004). Class noise vs. attribute noise: A quantitative study. *Artificial Intelligence Review*, *22*(3), 177–210.