Article:
Zhou, H., Wang, Y., Hancock, T. et al. (2026) Modelling the heterogeneity in preferences of subway passengers utilizing smart card data from Beijing. Transportation Research Part A: Policy and Practice, 205. 104887. ISSN: 0965-8564

https://doi.org/10.1016/j.tra.2026.104887

Modelling the heterogeneity in preferences of subway passengers utilizing smart card data from Beijing

Huiyu Zhou[a], Yanping Wang[a], Thomas Hancock[b], Charisma Choudhury[b], David Palma Araneda[b], MingHui Hou[a], Yacan Wang[a*]

[a] School of Economics and Management, Beijing Jiaotong University, Beijing, China

[b] Institute for Transport Studies & Choice Modelling Centre, University of Leeds, UK

Acknowledgment: Professor Yacan Wang was supported by the Natural Science Foundation of China [grant number 72474018]. Professor Huiyu Zhou was supported by National Social Science Foundation of China [grant number 22BJY211]. Professor Charisma Choudhury and Dr Thomas Hancock's time was supported by the UKRI Future Leader Fellowship [MRT020423/1]. We thank Beijing Mass Transit Railway Operation Corporation Limited for providing the anonymous data used for this research.The authors would also like to express their sincere gratitude to the anonymous reviewers for their constructive and insightful comments, which have greatly improved the quality of this manuscript.

Disclosure of potential Conflicts of Interest: None

*Corresponding Author: ycwang@bjtu.edu.cn

Modelling the heterogeneity in preferences of subway passengers utilizing smart card data from Beijing

Abstract: There is considerable heterogeneity in the travel behavior of subway passengers. The growing availability of large-scale smart card data offers the valuable opportunity to observe the major share of the passengers over a long period of time to uncover this heterogeneity in travel behavior and devise targeted policies for peak spreading. This study analyzes 7.92 million trip records from Beijing's subway system—spanning 24 lines and 331 stations—to identify distinct passenger groups based on travel frequency, temporal patterns, and spatial characteristics. Four distinct passenger clusters are identified, and class-specific mode choice models (between subway and bus) are estimated to derive each group's value of travel time (VOT) and value of travel time reliability (VOR). Results indicate that commuters who take the subway frequently have a higher willingness to pay to improve trip reliability. This indicates that measures like low fare incentives may not be effective in changing the travel behavior of this group, while they are the main target group for alleviating the subway congestion during peak-hours. Therefore, targeted incentives based on the travel characteristics and transport preferences of different passengers could better improve the effectiveness of subway congestion management.

Keywords: subway congestion management; transport attribute preferences; travel patterns; targeted incentives; smart card data

1 Introduction

With the large-scale development of urban rail transit, the subway or metro has become one of the most popular travel modes in China (W. Li et al., 2020). The densification of the urban population has led to a strong demand for travel, and the contradiction between supply and demand in the operation of the subway has become increasingly prominent. Many cities have severe overcrowding on their subway systems, which has a non-negligible impact on the reliability, safety and comfort of passengers' travel. Keeping up with growing demand by adding capacity is often difficult, especially in the short term (Halvorsen et al., 2020; Ma et al., 2020; Peer et al., 2016). Therefore, it is important to improve the efficiency of subway operations and alleviate the problem of congestion in the subway system during peak hours.

The literature on subway congestion management measures has focused on inducing subway users to stagger their trips through off-peak fare discounts, and generally adopts a "one-size-fits-all" policy. However, such measures based on departure times and financial costs do not take into account the impact of passengers' own attributes (e.g., travel purpose, transport preferences, etc.). Moreover, different types of passengers have heterogeneous sensitivity to the influencing factors, resulting in low response rates of passengers to a "one-size-fits-all" policy (Halvorsen et al., 2016). Although some studies involving travel mode choice considered the value of transport service attributes (e.g., travel time, reliability), they estimated willingness to pay mainly from stated preference data. A key limitation of stated preferences is that it can be subjected to *behavioural incongruence*, making the applicability of the findings in the real-world questionable (Dixit et al., 2017; Fifer et al., 2014). Much of the value attributed to punctuality stems from the stress caused to passengers by the potential to experience unexpected delays, which is difficult to convey in a stated survey (Bates et al., 2001). Fifer et al. (2014) also discussed the prevalence of the presence of hypothetical bias in stated choice experiments using comparisons of stated preference and revealed preference data, and propose some mitigation methods to minimize this bias.

In order to develop more efficient incentives, it is necessary to gain insight into the travel characteristics of public transport users and to explore the driving factors that can change their travel behavior in the real-world. However, it has been difficult to identify the travel characteristics and estimate the preferences of transport service attributes, which has become an obstacle to developing targeted incentives (Agarwal et al., 2020; Goulet-Langlois et al., 2016). With the widespread adoption of automated fare collection systems (AFC), continuous and passive collection of large data on mobility is possible. Smart card data is characterized by large scale, fast production and long investigation time, which provides data support for continuously tracking passengers' travel behavior and identifying passengers' travel patterns and behavioural characteristics. The use of smart card data thus helps us to more accurately and efficiently estimate their preferences and sensitivities towards different level-of-service attributes.

This paper aims to use large scale smart card data to integrate the travel characteristics and transport service attribute preferences of public transport users to develop targeted subway congestion management measures. The use of revealed preference data facilitates the accuracy and validity of the findings, as well as enabling long-term observation of travel behavior. The behavioral characteristics of passengers and their transport service attribute preferences are taken into account to develop targeted incentive schemes so as to coordinate the imbalance of supply and demand between the subway system and the bus system, as well as between different time periods. On the one hand, it provides supporting evidence for the study of the heterogeneity of public transport users' service attribute preferences, and on the other hand, it provides suggestions for the development of management policies to alleviate the subway peak-hour congestion.

The rest of this paper is organized as follows. Section 2 reviews the literature on transportation demand management and the value of transport service attributes. Section 3 presents the trip record data and user clustering results from the Beijing public transport AFC database, and describes the empirical analysis. Section 4 analyzes the results of the empirical study. Section 5 discusses the findings and proposes targeted incentives for different passenger groups. Section 6 summarizes the findings and presents the direction of future research.

2 Literature review

2.1 Transportation demand management

Traditionally, transportation demand management (TDM) has focused on road congestion and aims to influence users' travel behavior through intervention policies, mainly consisting of improving alternative travel modes, rewards (such as toll discounts, etc.) or penalties (such as congestion or road tolls, increased parking fees) (Gärling & Schuitema, 2007; Tillema et al., 2013). However, as more public transport agencies face congestion problems, cities have developed more structured concepts and methods for public transport demand management (PTDM). Examples include Melbourne's "early bird" free fare scheme (Currie, 2009), Hong Kong's route-based differentiated pricing and pre-peak fare discounts (Halvorsen et al., 2020; S.-M. Li & Wong, 1994), Singapore's INSINC scheme based on monetary and social incentives (Pluntke & Prabhakar, 2013). Other possibilities include random rewards (e.g., lotteries, raffles, etc.), employer incentives (e.g., tax incentives for employers who participate in 'flexible working time' system), provision of crowding information and so on (Henn et al., 2010). Studies of the implementation of these interventions reported that 2-5% of passengers participated in staggered trips to earn rewards, and that the potential for peak diffusion declined sharply as the time spent shifting to off-peak increased (Halvorsen et al., 2016; Henn et al., 2011; Ma & Koutsopoulos, 2019). San Francisco's BART (Bay Area Rapid Transit) Perks program added employer support for flexible work schedules, with about 10 percent of people who previously traveled during peak hours switching their rides to one of the shoulder or off-peak hours during the test period (Greene-Roesel et al., 2018). Subway users may be less motivated to change their habitual travel patterns, affected by personal characteristics, transport service attributes, work schedules, and other factors. Typical transport interventions using uniform discounts for a single time period, such as differential pricing and off-peak fare discounts, may actually be very inefficient with respect to the number of effective participants in the target population, with many users receiving rewards from the incentive program but not contributing to reducing peak period congestion (Halvorsen et al., 2016; Ma et al., 2020).

Differentiated incentives for different types of groups may be more effective. BART Perks Phase II offered individualized incentives to 1900 recruited BART passengers, and participants who received incentives for changing their departure time increased the percentage of trips taken during the incentive period by 6 to 20 percent (BART, 2019). Halvorsen et al. (2016, 2020) evaluated the effectiveness of a pre-peak fare discount strategy for the Hong Kong MTR. Passengers were divided into different groups based on the spatial and temporal characteristics of historical trips, identifying passengers who changed their behavior and analyzing the main factors that contributed to this change. The results show that different groups respond heterogeneously to promotions, and that the shift time required for passengers to participate in a discount, the variability of departure time, and price sensitivity are the main influencing factors. Wang et al. (2020) used smart card data to segment passengers to identify how different types of user groups respond to the discount policy in the context of the Beijing subway staggered discount policy, and indicated that fare discount measures are more effective for passengers with high price sensitivity. Wang et al. (2020) only measured the heterogeneous responses of different groups to the staggered discount fare policy implemented in Beijing in 2016, and did not estimate the effect of travel preferences. Zou et al. (2019) revealed that retiming elasticities differed significantly across passenger groups, with low-frequency passengers being more sensitive to fare discounts than high-frequency passengers, and the elasticity decreasing significantly with increasing time shifted from the previous peak departure time to the off-peak.

Henn et al. (2010, 2011) reviewed and evaluated a range of demand management policies and instruments for urban rail transit and other industries which experience similar peak challenges, in addition to a survey analysis of the willingness of Sydney commuters to change their trips before the peak in order to gain incentives. The results show that targeting specific user groups, combining different types of interventions, and integrating policies between transport providers and modes can make demand management programs more effective.

The widespread deployment of AFC has provided very valuable material for TDM research. As a result of technological advances, continuous and passive collection of large amounts of data on mobility is possible with little or no burden on the respondent. Smart card data provides larger sample sizes and longer survey periods than traditional survey data, allowing demand patterns to be studied at a more granular level (Ali et al., 2015; Bagchi & White, 2005). Extracting travel patterns using smart card data is critical for inferring trip purpose or passenger attributes (e.g., price sensitivity, etc.), helping to continuously monitor changes in passenger behavior and enhance understanding of passenger behavior (Alsger et al., 2018; Kusakabe & Asakura, 2014). For privacy protection purposes, there is a lack of socio-demographic variables in smart card data, thus customer segmentation is an appropriate means to compensate for this deficiency. Monitoring the travel characteristics of different user groups contributes to the design of more effective strategies to take heterogeneous behavior into account (Ma et al., 2020).

2.2 The impact of transport service attributes on travel decisions

Identifying travel characteristics and exploring the driving factors that change passengers' travel behavior are crucial to developing targeted interventions. Previous researches on subway congestion management have focused more on the trade-off between departure time and monetary costs, neglecting the impact of service reliability on travel decisions (Agarwal et al., 2020; Carrion & Levinson, 2012; van Loon et al., 2011). Travel time and travel time reliability are two major factors influencing travel choices, and many studies have focused on the heterogeneous preferences of motorists for road service quality choices. Assessing the impact of travel time, reliability, and monetary costs through drivers’ choice decisions between free and congestion-based toll roads in the experiment provides valuable insights into toll pricing schemes for drivers’ different traffic preferences (Lam & Small, 2001; Liu et al., 2004, 2007; Small et al., 2005). However, there are relatively few studies on the preferences of public transport users. Ignoring travel time reliability may bias the assessment of strategies aimed at improving the efficiency of urban transportation systems and may lead to inappropriate policy actions (Agarwal et al., 2020; Bhat & Sardesai, 2006; Carrion & Levinson, 2012; Zhu et al., 2017). Asensio and Matas (2008) conducted a stated preference experiment to capture the impact of travel time variability on the choice process through the value of schedule delay early and the value of schedule delay late by commuters in a metropolitan environment. Their results show that the value of schedule delay late is on average 2.4 times the value of travel time. Agarwal et al. (2020) was the first who used smart card data to evaluate the travel preferences of public transport passengers by estimating the value of travel time (VOT) and value of reliability (VOR). The results show that there is significant heterogeneity in the transport preferences of different types of passengers, and that travel time and reliability have important effects on commuters' travel choices.

2.3 Review

From the literature review, it can be concluded that "one-size-fits-all" travel interventions ignore the heterogeneity among passengers and may be less effective. The results of various empirical researches prove that there is significant heterogeneity in VOT and VOR for different types of passengers and that they have important effects on travel choices. However, the literature on subway peak-staggering policies contains less analysis of VOT and VOR, and the existing studies involving the value of travel time savings and unreliability primarily adopt stated preference data (Bates et al., 2001; Hollander, 2006). Since stated choice surveys rely heavily on respondents' subjective judgments, there is likely to be a discrepancy between respondents' stated preferences and actual actions taken in such scenarios, leading to noise and bias in responses (Dixit et al., 2017). Additionally, much of the value attributed to punctuality stems from the stress caused to passengers by the potential to experience unexpected delays, which is difficult to convey in a stated survey (Bates et al., 2001). Agarwal et al. (2020) proposed a model for computing VOT and VOR for public transport users and conducted an empirical analysis using Singapore smart card data, but did not analyze the travel characteristics of users in detail. Additionally, segmenting users based on socio-demographic variables (e.g., elderly, adults, students) assumes that travel characteristics and preferences of people in the same demographic group or life stage are homogeneous, which may be inaccurate.

The work in this paper avoids the shortcomings outlined above by using large smart card transaction data to capture passengers' travel patterns. It models the choice between two modes: subway and bus, which have very different travel times, reliability and cost in the same travel corridor. This enables us to estimate the VOT and VOR for different passenger groups. It is assumed that there are also significant differences in travel characteristics among commuters. For example, some passengers rely heavily on the subway for commuting, while some only take the subway once a day to or from work, at which point their preferences for transport service attributes are also distinct. Meanwhile, transport service attribute preferences are also closely related to periodic travel schedules (Tseng & Verhoef, 2008), indicating that the analysis of travel characteristics is necessary for passengers' travel decision studies.

3 Data and methods

This section describes the smart card dataset used, the clustering algorithm used to analyze the travel characteristics of different categories of travelers, and the choice models used to capture the heterogeneous preferences of the travelers.

3.1 Beijing smart card data processing

Beijing, the capital of China, has a resident population of 21,893,000 in 2020 (Source: Beijing Municipal Bureau of Statistics). The Beijing subway has always been one of the largest and most congested transportation systems in China (Zou et al., 2019). Beijing's public transportation, including the bus system and subway system, fully implements automatic fare collection. Passengers can use the smart card to pay the fare. During each transaction of the smart card, the card ID, transportation mode (subway/bus), check-in time, check-in station, check-out time, check-out station, line number, and station latitude and longitude are recorded. The fare structure of the public transport in Beijing during the survey period is shown in Table 1. Smart card users can enjoy a 50% discount on the bus, and the fare is much lower than that of the subway. Beijing public transport smart card system transaction records for the whole month of November 2019 are used here to study the travel characteristics and travel preferences of passengers. The data includes a total of 185,135,824 trips.. Figure 1 shows the study area of this paper, covering 331 subway stations and

24 lines.

More specifically, the Beijing CBD is an urban functional area with the development of service industries as its core, and the leading industries are finance, insurance, telecommunication, information, services and consultancy, situated in the middle of Chaoyang District in a 7 square kilometre rectangular area between Dongdaqiao Road, East Fourth Ring Road, Tonghui River and Chaoyang North Road.

Table 1. Fare structure of public transportation in Beijing (Exchange rate using June 2024 data).

Subway			Bus		
Mileage (km)	Ticket Price (yuan)	Equivalent in US$	Mileage (km)	Ticket Price (yuan)	Equivalent in US$
0-6	3	0.41	0-10	1	0.14
6-12	4	0.55	10-15	1.5	0.21
12-22	5	0.69	15-20	2	0.28
22-32	6	0.83	20-25	2.5	0.35
32-52	7	0.97	25-30	3	0.41

Note: Subway fares increase by 1 yuan for every 20 kilometers over 32 kilometers; bus fares increase by 0.5 yuan for every 5 kilometers over 10 kilometers.

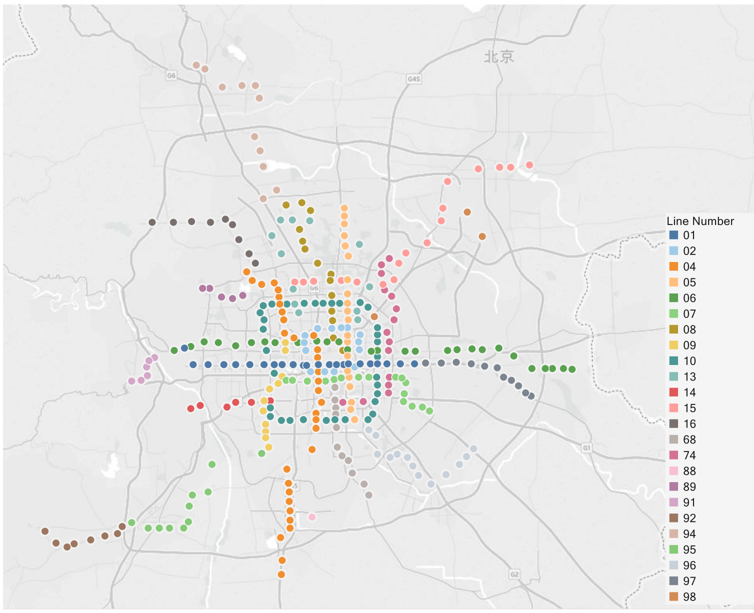


Figure 1. Map of the study line

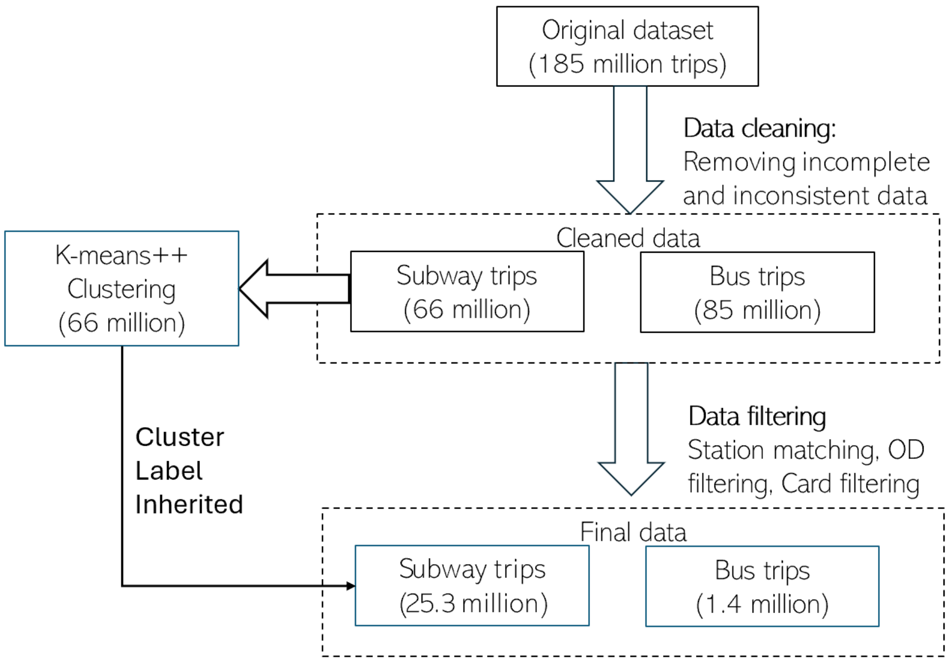


Figure 2. Data cleaning and filtering process

The data cleaning and filtering process is illustrated in Figure 2, the original dataset, including both subway and bus trips, included 185 million records. The dataset was cleaned first, where the incomplete trips (e.g. where the origin or destination stop was not recorded due to some error) and trips with spatial or temporal inconsistencies (e.g. trips with duration shorter than 60 seconds, implausible geographical coordinates, etc.) were removed. This ensured the reliability of subsequent analyses and resulted in a cleaned dataset consisting of 66 million subway trips and 85 million bus trips (151 million trips in total).

Given the focus is on multimodal analyses, only trips where both bus and subway were available alternatives to the travelers were then filtered out. The following three-step process was followed in this regard:

1. Station/Stop filtering: Bus stations/stops which are not located within 300m of a subway station were removed. This approach ensures consistent accessibility standards by considering only bus stops within 300 metres of subway stations as valid matches, regardless of local infrastructure density. This mitigates geographic selection bias inherent in transit network design. Since all subway stations had at least one bus stop within 300m, this did not lead to any reduction in the number of subway stations.

2. OD filtering: Only ODs which were served by at least one subway and one bus line were retained.

3. Card filtering: Only cards of users whose records of using either both subway and bus or only subway over the one-month period were retained. Note that we retain subway trips from users who did not utilize bus alternatives, provided their trips occurred along corridors where bus options were objectively available (within 300m of both origin and destination stations). This spatial matching criterion ensures that both alternatives are available in the trips analyzed.

As shown in Figure 2, the cleaned dataset separates into subway (66 million trips) and bus (85 million trips) branches for multimodal analysis. For subway trips, we first apply clustering to identify traveler subgroups before performing Origin-Destination (OD) matching, enabling cluster-specific estimation of time and reliability values (VOT/VOR) later. Bus trips are integrated then through direct matching, where we link bus stops located within 300m of subway stations (see also Figure 2). The difference in matched volumes primarily stems from the spatial precision requirements when connecting these distinct transit systems.

In order to capture the heterogeneity, the subway trips were grouped using clustering techniques. k-means++ clustering was used in the regard to divide the users to four clusters so that cluster-specific values of time (VOT) and reliability (VOR) can be determined for each passenger segment. For details of the clustering process, please refer to Section 3.2.

The clustering was intentionally performed prior to any OD matching using the full 66 million observations. This approach was designed to first identify traveler segments based on comprehensive behavioral patterns across the entire dataset, ensuring an objective and unbiased grouping. Subsequently, we examine mode choice behavior within the context of available alternatives for each pre-defined group. Since the 25.3 million OD-matched trips are a subset of the original 66 million, every card retains its pre-assigned cluster label, guaranteeing consistency between the segmentation and the subsequent discrete choice modeling.

Following the identification of four traveler clusters through k-means++, we analyze subway trips within each cluster to estimate mode-specific values of time (VOT) and reliability (VOR). For every origin-destination (OD) station pair in each cluster, we first aggregate all corresponding subway trips. Using ArcGIS, we then identify matching bus trips where both origin and destination bus stops fall within 300m buffers of the OD pairs' corresponding subway stations (as demonstrated in Figure 2). This matching process serves two critical purposes: (1) ensuring commuters had genuine modal alternatives, and (2) enabling direct comparison of travel time valuations between modes.

Within each cluster, trips without matched bus alternatives—indicating passengers only had access to the subway and lacked feasible bus options—are excluded from estimation analysis. Given we aim to understand the preferences of subway users, we further filter bus observations to remove trips made by bus-only travelers who never use the subway.

This process yields our final dataset of approximately 25.3 million subway trips (by approximately 4.89 million cards) with comparable 1.4 million bus trips (by approximately 0.96 million cards, a subset of the 4.89 subway users), distributed across different behavioral clusters.

These three filtering steps ensured commuters had genuine modal alternatives and enabled direct comparison of travel time valuations between modes. It may be noted that among these matched users, 61.7% (4.89 million cards) had trips in OD pairs containing both subway and bus options (qualifying as multimodal candidates), though only 19.6% of them (0.96 million cards) were observed to actually use both modes. The remaining users used only subway in matched OD pairs (3.93 million cards). Due to the risk of introducing potential biases, all travelers (regardless of if s/he chose only one mode or both) have been included in the analyses.

Note that we retain subway trips from users who did not utilize bus alternatives. This was because, the OD filtering process ensured that the subway trips occurred along corridors where bus options were objectively available (within 300m of both origin and destination stations) and it was a choice not to use the bus.

Therefore, the trips without alternatives are removed from the dataset, only matched subway trips and bus trips remained in the dataset for VOT and VOR calculation.

This process also eliminates data from users whose subway and bus trips do not match, making it easier to estimate the continuous mode choice behaviour of passengers. After the filtering process, we retain 26,791,973 trips made by 4,892,551 cards for analysis of passengers' travel preferences. Due to the risk of introducing potential biases, all travelers (regardless of if s/he chose only one mode or both) have been included in the analyses.

As shown in Appendix Table A.1, the final dataset includes 4.89 million subway users (cards) whose trips occurred in OD pairs with bus alternatives (25.3 million subway trips). The 4.89 million cards represent users who had both subway and bus options available in their OD pairs. Of these individuals, only 0.96 million cards (19.6% of 4.89M) were observed to actually use bus services (generating 1.4 million bus trips), while the remaining 3.93 million cards (80.4%) used subway exclusively despite available bus alternatives. The study excluded 5.56 million bus-only users (30.2 million bus trips) who never used subway, as they fall outside the scope of subway users' mode choice analysis. This data structure effectively isolates the population for studying subway users' mode retention and switching behavior when alternative options exist.

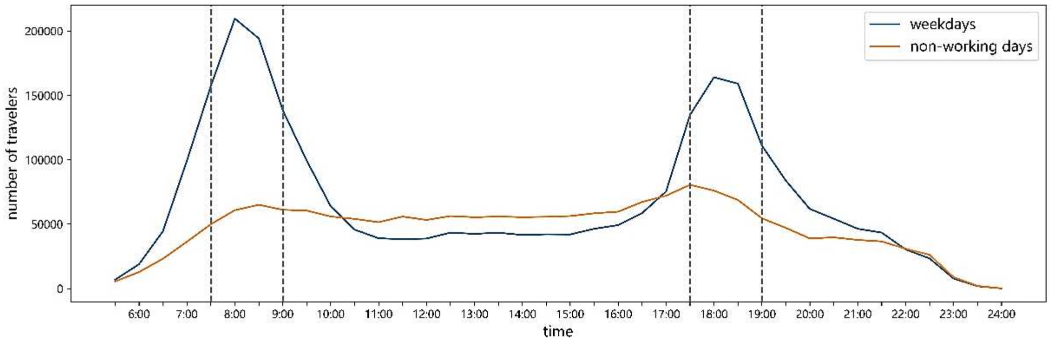


Figure 3. Distribution of the average number of travelers.

According to the average number of system-wide subway trips on weekdays and non-working days, respectively, we checked the distribution of passenger volume during a day. As shown in Figure 3, there is a sharp increase in the number of trips from 7:30 to 9:00 and 17:30 to 19:00 on weekdays, with very high peak values, and the morning peak is more congested than the evening peak. On non-working days, the highest number of trips is observed around 8:30 and 17:30, but the variation in demand levels is not very significant compared to weekdays. Therefore, it is determined that the morning peak time period is from 7:30 to 9:00 and the evening peak time period is from 17:30 to 19:00.

3.2 User clustering

The smart card database includes detailed information about almost all passengers' trips over one month, making it possible to extract the travel patterns of passengers. The objective of the clustering was to segment passengers based on general behavioral characteristics and identify groups who are more likely to have similar responses to a specific intervention.

Clustering was performed using the November 2019 Beijing subway transaction record data. The dataset includes more than 66 million trips completed by approximately 7.92 million passengers. It is assumed that each card represents an individual passenger. According to Beijing's smartcard setting, the smartcard is tied to a person's mobile phone and identity document, and it is generally not possible to lend it to someone else. As you must swipe your smartcard every time you enter or leave a subway station to complete a journey, the system does not allow you to swipe your smartcard twice or more, or to swipe it for someone else, ensuring that one journey corresponds to one card and one person. If someone else wants to use the Tube temporarily, they will need to buy their own temporary ticket, which is outside the scope of our smartcard data.

Due to the large amount of data, the k-means++ algorithm (Arthur & Vassilvitskii, 2007) was chosen for this clustering analysis, after taking into account the computational speed and the accuracy of the results. This algorithm selects clustering centers that are more distantly distributed, which improves the randomness of the initial assignment of the k-means algorithm. It reduces the time for clustering convergence, improves the accuracy, and is more conducive to ensuring the global optimum (Arthur & Vassilvitskii, 2007).

Referring to the study of Halvorsen et al. (2020) on public transport demand management, the clustering variables used to capture the travel characteristics of passengers in terms of three dimensions: frequency of travel, temporal characteristics and spatial characteristics are set as follows (Table 2). As there are significant differences in travel behaviour between weekdays and non-working days, different values are calculated for weekdays and non-working days, except for the variables gap_times, gap_mean, mor_peak_num and eve_peak_num. To understand the meaning of gap, it is useful to first define the number of days between two consecutive trips without a trip as gap_days, e.g. if there is a trip on 1 November and the next trip date is 4 November, then gap_days is 2. Then gap_mean is the average of the gap_days and gap_times is the number of times the gap was taken. As there is no significant peak on non-working days, mor_peak_num and eve_peak_num only count weekday peak trips.

Table 2. Clustering variables definition.

	Variables	Describe
Travel frequency	days	Total number of travel days
	trips	Median number of trips in a day
	gap_times	Number of intervals (of at least one day without travel)
	gap_mean	Mean value of travel interval length (days)
Temporal characteristics	first_trip	Median time of day to start the first trip
	first_dura	Median travel time of the first trip of the day
	last_trip	Median time of day to start the last trip
	first_last_interval	Duration of the interval between the first and last trip of the day
	mor_peak_num	Number of trips during the morning peak (7:30-9:00) on weekdays
	eve_peak_num	Number of trips during the evening peak (17:30-19:00) on weekdays
Spatial characteristics	ODs	Number of unique OD (origination - destination) pairs travelled
	en_longitude, en_latitude	Latitude and longitude of the origination with the highest frequency of first trips
	ex_longtitude, ex_latitude	Latitude and longitude of the destination with the highest frequency of first trips

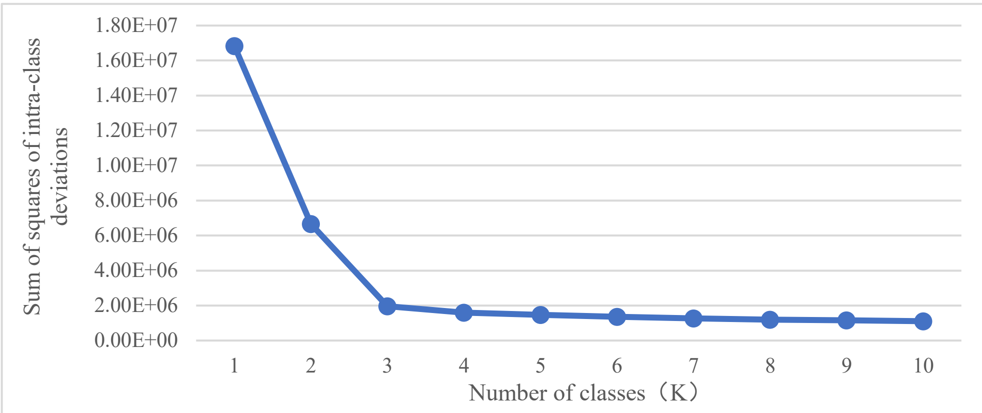


Figure 4. Variation of the sum of squares of intra-class deviations with the value of K.

The use of the k-means++ algorithm requires that the number of clusters be specified first. The optimal number of clusters can be determined by statistical metrics, but interpretability and meaningfulness of the results are also worth considering. We use the "elbow" method (Nainggolan et al., 2019) which selects the number of clusters (K) based on when the maximum reduction in the sum of squares of intra-class deviations occurs to determine the number of clusters. Data were standardized before clustering using min-max normalization method before clustering. Figure 4 shows the values of the sum of squares of the intra-class deviations for K=1 to K=10 and reveals that the intra-class deviations do not decrease substantially beyond 3 clusters.

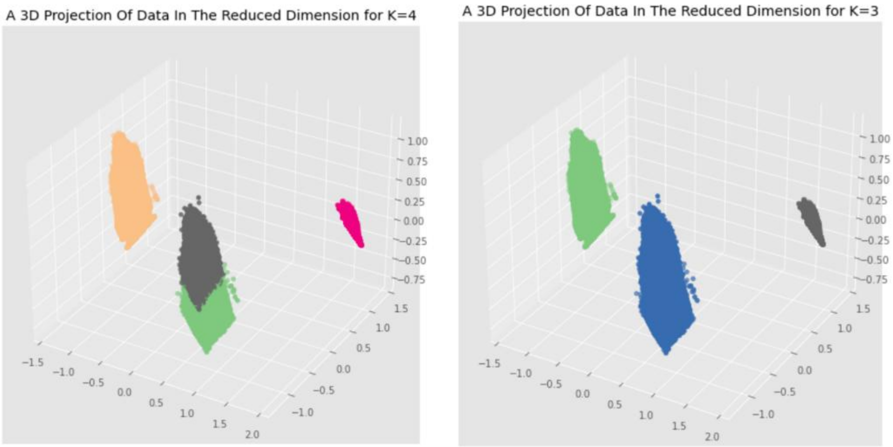


Figure 5. Distribution of clustered groups by PCA dimensionality reduction for K=4 and K=3.

We reviewed the features of the clusters for K ⩾ 3 and found that K = 4 captured more additional behavioral features, while clusters larger than 4 provided little further benefit. As shown in Figure 5, the distribution of clustered groups by Principal Component Analysis (PCA) shows that the classifier clusters Group 3 and Group 4 into one group when K = 3 is chosen. These two groups have similar travel patterns, while there is a significant difference between them and the other two groups, which is the reason why the clustering evaluation index considers it optimal to classify them into 3 groups. However, considering the practical meaning of specific variables, Group 3 and Group 4 have a large gap in major characteristics, as shown in Table 4.

For example, in terms of the time of the first trip on a weekday, Group 3 (9:05) is earlier than Group 4 (13:11) by an average of 4 hours, and in terms of the number of days of weekday trips on the subway in a month, Group 3 (16.30 days) is 4 times higher than Group 4 (3.18 days), and in the number of trips taken during the morning peak period, Group 3 (3.28) is twice as frequent as Group 4 (1.33). These differences in departure times and frequency are important for the precise implementation of subway congestion interventions, hence the choice to categorize the population into 4 groups.

Furthermore, a bootstrap scheme was run to assess the stability of the clustering using the R-package FPC and the results show that the Jaccard coefficient for all four groups is greater than 0.85 as shown in Table 3, indicating that the clusters are highly stable (Hennig, 2007).

Table 3. Clustering stability results for the bootstrap scheme

	Group 1	Group 2	Group 3	Group 4
Jaccard coefficient	1.00	0.99	0.97	0.93

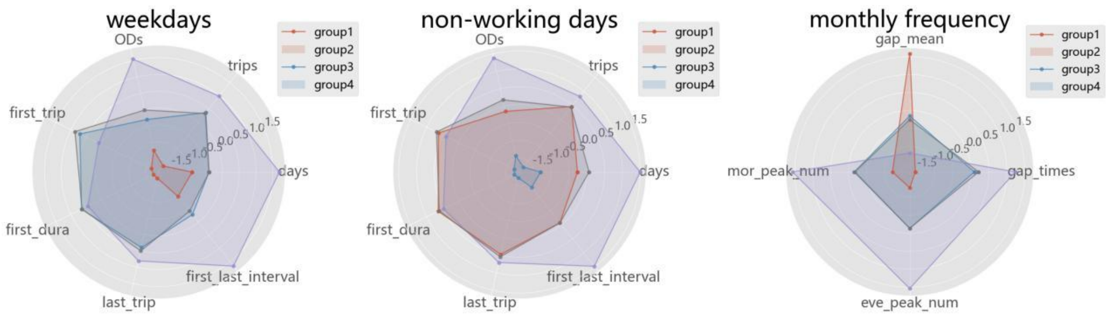


Figure 6. Radar chart of the means of each group on the major features.

Table 4 Mean values of each group on the major features.

	Variables	Group1	Group2	Group3	Group4
Weekdays	Number of trips per day	0.00	1.38	1.81	1.36
	Travel days per month	0.00	3.14	16.30	3.18
	Number of ODs	0.00	2.49	7.37	3.26
	Time of the first trip in day	--	12:21	9:05	13:11
	Duration of the first trip in day	--	0:43	0:39	0:43
	Time of the last trip in day	--	14:29	17:16	15:11
	Interval between the first and last trip	--	1:53	7:18	1:31
	Number of trips in morning peak	0.00	1.35	3.28	1.33
	Number of trips in evening peak	0.00	1.12	2.53	1.11
Non-working days	Number of trips per day	1.43	0.00	1.67	1.41
	Travel days per month	1.38	0.00	3.74	1.82
	Number of ODs	1.84	0.00	4.05	2.31
	Time of the first trip in day	12:54	--	11:38	13:15
	Duration of the first trip in day	0:45	--	0:41	0:44
	Time of the last trip in day	14:49	--	16:23	15:17
	Interval between the first and last trip	1:51	--	4:08	1:51

Note: -- represent not available

Figure 6 and Table 4 shows a comparison of the mean values for each cluster on several primary characteristics. According to these characteristics, the key features of travel patterns of the four clusters were identified as follows:

Group 1- Leisure travelers: These travelers rarely take the subway on weekdays but do travel by subway on non-working days. They have a small number of trip intervals, but a long duration of each interval, meaning that they travel very infrequently by subway, usually with longer intervals between trips. This implies that they mainly take the subway for leisure travel.

Group 2- Low-frequency travelers: These travelers occasionally travel by subway on weekdays and travel between relatively regular OD pairs, meaning they frequently go from and to the same locations. They tend to travel later in the day, less often during the morning and evening peak hours, but for a longer travel time. It is likely that these travelers have already adopted the subway staggered travel behavior on their own. On non-working days, they rarely take the subway. The frequency of trips by subway is relatively low, such that the interval between trips is long.

Group 3 - High-frequency travelers: This group takes the subway most frequently and has a high demand or habit of using the subway on both weekdays and non-working days, but for a shorter travel time. They often take the subway during morning and evening peak hours and have short gaps between trips. Therefore, they have the highest loyalty in subway usage. It is presumed that they are likely workers who mainly use the subway for commuting and are thus the major target group of the staggered travel policy.

Group 4 - Intermittent travelers: The travel characteristics of this group on weekdays are similar to those of Group 2, but the difference is that they tend to travel by subway on non-working days. They start their first trips relatively late in the day and make fewer trips during peak periods, but the locations are more diverse than Group 2. This group shows a higher *stickiness* on subway transport than Group 2. They may also have adopted the subway staggered travel behavior on their own during weekdays. However, they are more dependent on the subway than Group 2 and will switch back to subway travel habits on non-working days.

Table 5. Distribution of cards and trips in each group.

	trips		cards	
	number	percentage	number	percentage
Group1	2,829,278	4.26%	1,403,681	17.72%
Group2	16,660,181	25.09%	3,485,208	44.00%
Group3	31,419,490	47.32%	863,633	10.90%
Group4	15,485,594	23.32%	2,167,700	27.37%

The percentage of individuals and trips for each group is shown in Table 5. Group 3 had the lowest number of individuals at 10.90%, but completed the most trips at 47.32%. Group 2 had the highest number of individuals at 44.00% and completed 25.09% of the trips. Group 1 had the lowest percentage of trips at 4.26%.

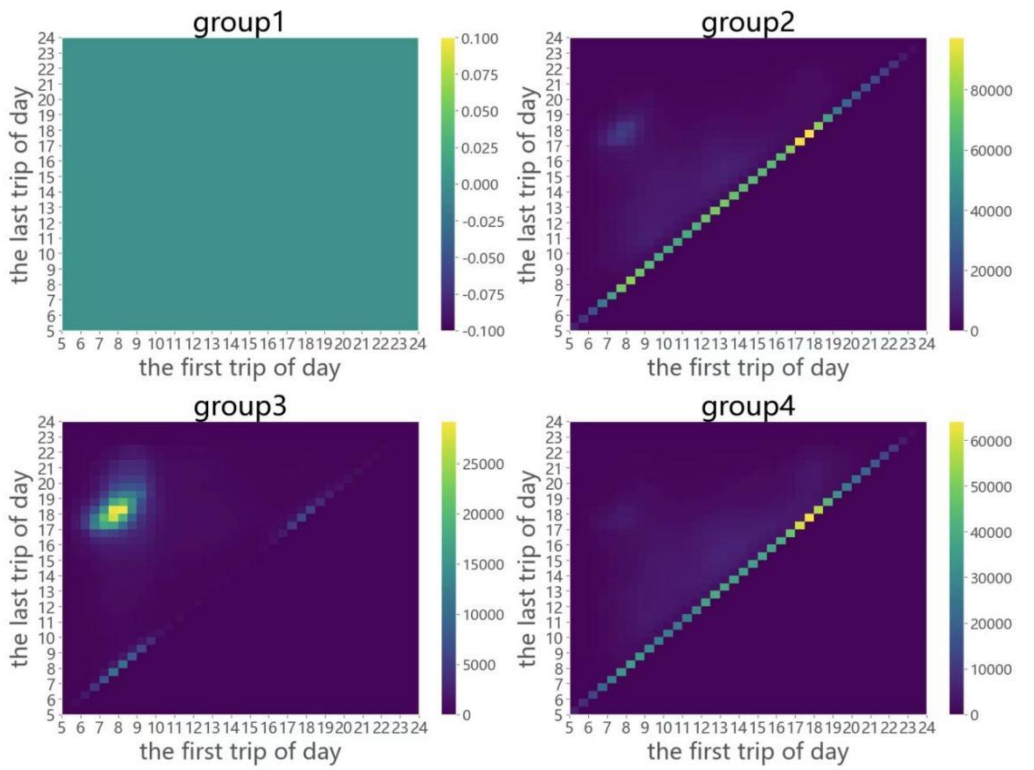


Figure 7. Distribution of the first trip time and the last trip time of day on weekdays for each group

The distribution of the median time of the first and last trip of the day on weekdays for each group is shown in Figure 7. The horizontal coordinate of the graph indicates the time of the first subway trip of the day, the vertical coordinate indicates the time of the last subway trip of the day, and the colour of the blocks reflects the relative size of the number of passengers at a given first and last entry time into the subway station, with the yellow blocks representing common combinations of times when ridership peaks. The combination on the diagonal indicates that the first and last entry times are the same, i.e. there is only one trip on the subway per day.

The first group of passengers rarely travels by subway on weekdays, so the color change is not visible in the graph. This group is likely to have other alternative modes of transportation or travel less in general, and therefore have a low demand to travel by subway on weekdays. In fact, among the most frequent departure places in Group 1, subway stations near college POI points ranked high, which further verifies that this group includes a large number of probable college students who are busy with their courses at school on weekdays and take the subway to shopping areas and tourist attractions with the purpose of leisure on non-working days. From the graphs of Group 2 and Group 4, the color blocks on the diagonal are relatively darker, indicating that most people in these two groups take the subway only once in a day. Both groups reach their peak number of trips around

18:00, indicating that they only take the subway when they go home in the evening and do not travel by subway in the morning. Group 2 has a small number of passengers whose trip characteristics match the commuting pattern, with the first trip in the morning peak (7:30-9:00) and the last trip in the evening peak (17:30-19:00). A large number of trips in Group 3 correspond to the commuting pattern, with peaks occurring at a combination of times when the first trip is during the morning peak and the last trip is during the evening peak. A small percentage of passengers in Group 3 take the subway only once in the morning or in the evening, but their trips are close to work and off work hours, respectively. It can be inferred that almost all of Group 3 are commuters relying on travelling by subway. They are the largest target group for subway congestion during peak hours. This group accounts for 10.9% of the total number of passengers in this study and needs to be focused on targeted interventions.

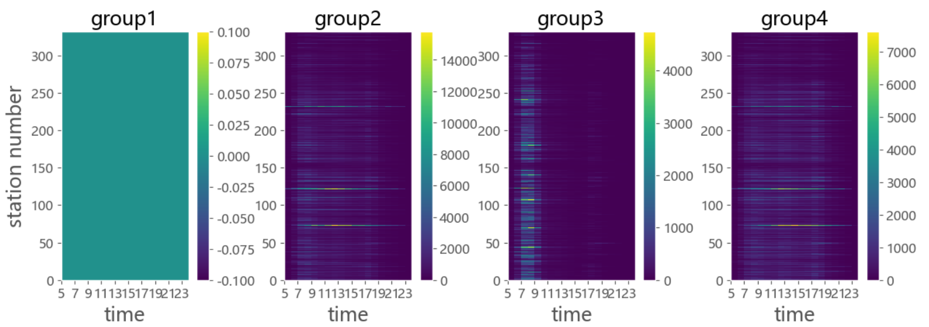
Figure 8 shows the distribution of the most frequent origin and destination for the first trip of each group on weekdays, with the stations listed in the order of their numerical numbers. Table 6 shows the top three stations of origin and destination for each group, where high-speed rail stations include Beijing Station, Beijing South Station and Beijing West Station, through which passengers can transfer to high-speed rail for cross-provincial travel.

Figure 9 labels the geographic locations of the subway stops mentioned in Table 8. It can be observed that the most frequently visited subway stations are distributed relatively evenly, with the majority situated within Beijing's 5^{th} ring road, with the exception of Tiantongyuan Bei(N) and Tiantongyuan stations, which are located in the Tiantongyuan neighbourhood[1] between Beijing's 5^{th} and 6^{th} ring road. The Tiantongyuan neighbourhood, situated in the north of Beijing, is the largest residential area in Asia, with a population exceeding 1 million. It is also home to 5 subway stations. This provides an explanation as to why the two stations depicted in Figure 9 are among the most frequently visited subway stations in Beijing.

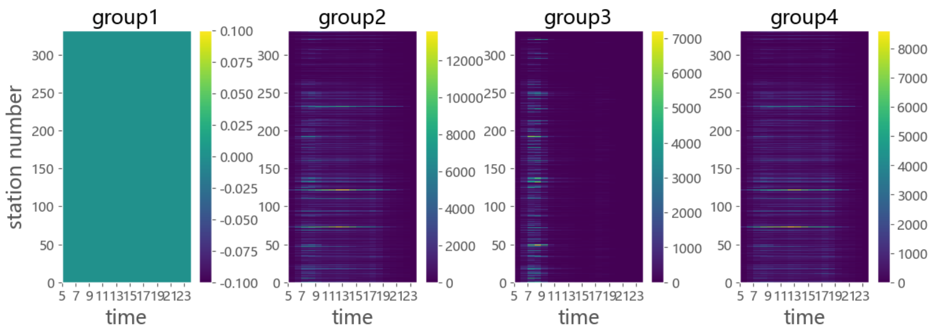
As shown in Figure 8 and Figure 9, Group 1 rarely travels by subway on weekdays, but travels by subway to tourist attractions (Xidan Station, Tian'anmendong Station) on non-working days, mainly for leisure trips (Table 6). The check-in and check-out stations for the first trip of the day in Group 2 and Group 4 are relatively concentrated, with fewer primary stations than in Group 3. These

[1] https://www.chinadaily.com.cn/a/201902/12/WS5c620fb1a3106c65c34e8cc4.html

two groups travel relatively late and have a large number of passengers coming from or going to high-speed rail stations, indicating that they are more likely to be passengers who need to make cross-provincial trips for reasons such as business or tourism, and therefore have relatively flexible schedules. The analysis of travel characteristics shows that the main difference between Group 2 and Group 4 is that Group 2 has very few trips on non-working days, while Group 4 has more records of taking the subway on non-working days, suggesting that Group 2 is more likely to be passengers who come to Beijing for short-term business or tourism. The first trip of the day for the third group is almost always concentrated between 7:00 and 9:00, with a wide distribution of check-in stations and relatively centralized check-out stations near the POI of the Beijing Central Business District (Chaoyang Men Station, Guomao Station), which is very consistent with the travel characteristics of commuters.



(a)



(b)

Figure 8. Distribution of the number of people in each group at the station where their first trip is most frequent on weekdays, (a) origination station, (b) destination station

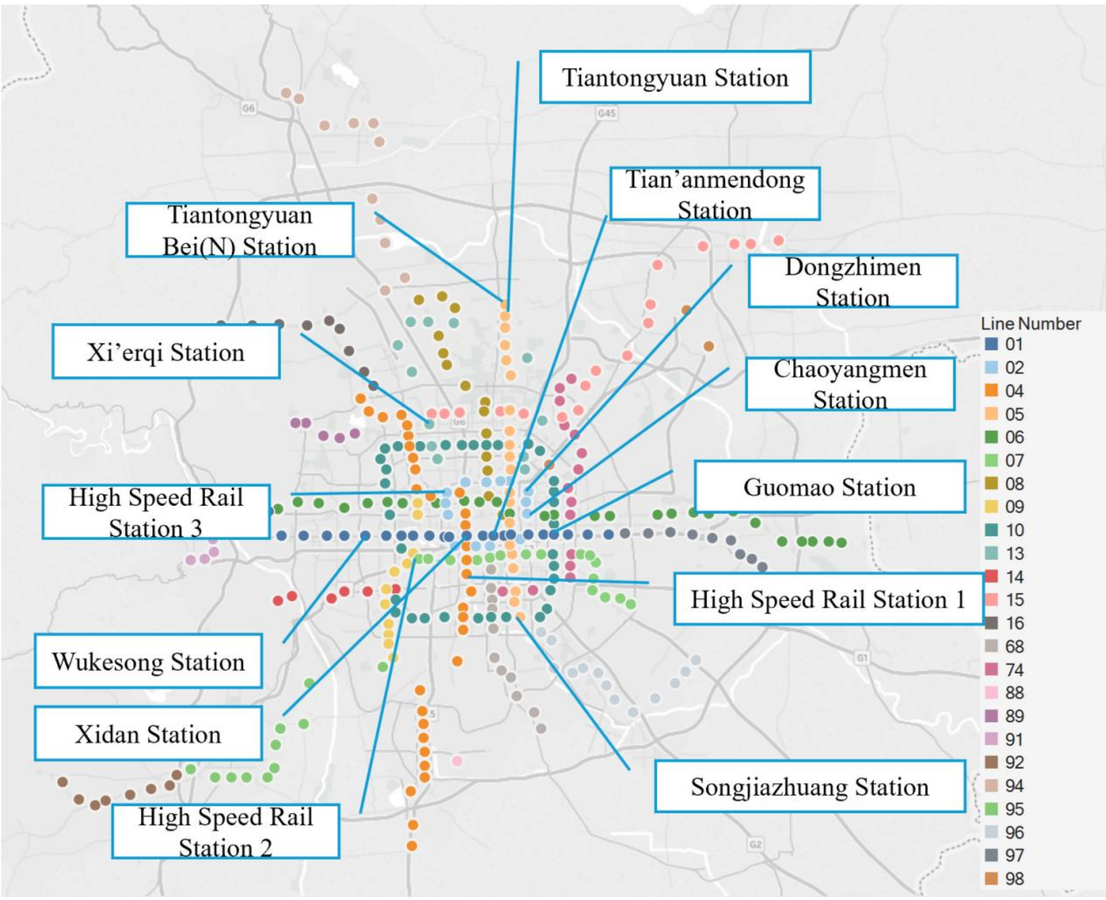


Figure 9. The location distribution of the most visited stations

Table 6. The 3 most frequent origination and destination stations for each group

	Group 1	Group 2	Group 3	Group 4
Origination station	high-speed rail stations, Dongzhimen Station, Tiangong Yuan Station	high-speed rail stations, Dongzhimen Station, Guomao Station	Songjiazhuang Station, Tiantong Yuan Station, Tiantongyuan Bei (N) Station	high-speed rail stations, Dongzhimen Station, Guomao Station
Destination station	high-speed rail stations, Xidan Station, Tian'anmendong Station	high-speed rail stations, Dongzhimen Station, Wukesong Station	Xi'erqi Station, Chaoyang Men Station, Guomao Station	high-speed rail stations, Dongzhimen Station, Wukesong Station

The trips starting and ending from CBD were filtered separately, and the proportion of the number of trips in each group is shown in Table 7. There are 3,347,114 trips related to the CBD, 1,715,891 start in the CBD and 1,749,449 end in the CBD, of which 118,226 start and end in the CBD. Among the trips related to CBD, Group 3 accounted for the highest number of trips with 49.21%, while Group 1 accounted for the least with only 2.85%. There are a small number of trips in Group 2 and Group 4 that start or end in CBD, accounting for 28.04% and 19.90%, respectively.

Meanwhile, groups 2 and 4 traveled slightly more from the CBD than they did to the CBD. There are 13.13% and 10.58% of trips in Group 2 and Group 4 are related to CBD, respectively, and we assume that the trip purpose of this segment is work, i.e., commuters. Combined with the finding in Figure 7 that the darker color blocks in Group 2 and Group 4 occur during the evening peak period (17:30 to 19:00) and are on the diagonal, it is inferred that this segment of commuters takes other modes of transportation to work and take the subway to get off from work.

Table 7. Proportion of CBD-related trips in each group (%)

	Group 1	Group 2	Group 3	Group 4
With the CBD as the departure point	2.93	28.24	48.52	20.31
With the CBD as the destination	2.75	27.81	49.75	19.69
All trips (Entering or leaving the CBD)	2.85	28.04	49.21	19.90

3.3 Empirical analysis measurement

In order to further analyze each group's preferences for transportation service attributes to guide the design of targeted incentives, we estimated VOT and VOR for different groups by referring to the theoretical model of Agarwal et al. (2020).

Passengers typically make their transportation mode choice before their trips begin. During the study period, the travelers were able to view (in real-time) the estimated time it would take for buses and subways to get to their destinations, the estimated arrival time of the next bus, and the corresponding fare information via the Internet and mobile phones. Therefore, it is assumed that public transport users are aware of the estimated travel time and travel cost of each mode of transportation for the trip from the origination to the destination prior to departure. The final origin/destination to the station/stop (and hence the access and egress time) are unobserved in our data. However, in the context of China, the boarding/access and alighting/egress time within the subway stations are also often non-trivial. For each subway trip, the within station/stop access time is calculated as the duration between the card tap-in and the first recorded train departure at the origin station, while egress time reflects the interval between the last train arrival and tap-out at the destination station. These are taken into account while calculating the travel times. Then, for a given OD, the expected travel time and travel time unreliability are calculated from the mean and standard deviation of multiple trips for the same OD, and the expected travel times for bus and subway are calculated from bus trips and subway trips, respectively. In addition, there can be large differences in travel times between weekdays and non-weekdays and between peak and off-peak periods, so expected travel times are calculated separately for these four periods.

Since buses travel on roads with more complicated road conditions and they travel at a lower speed than the subway, we assume that the travel time from the same origination to the destination is longer for the bus than the subway, and the unreliability of the travel time is higher as well. However, according to the fare structure in Beijing, the bus fare is lower than the subway. Hence, there are trade-offs for travel time, travel time reliability, and cost when passengers choose a mode of transportation.

The mode choice model for public transport can be formulated as follows:

$$U_{in}^{subway} = 0$$

$$U_{in}^{bus} = \beta_0 + \beta_{tt}\Delta(travel\ time)_{in} + \beta_{tr}\Delta(TTU)_{in} + \beta_{tc}\Delta(travel\ cost)_{in} + \beta_{CBDO}\text{CBDO}_{in} + \beta_{CBDD}\text{CBDD}_{in} + \beta_{CBDO1}\text{CBDO}_{in} \times \Delta(TTU)_{in} + \beta_{CBDD1}\text{CBDD}_{in} \times \Delta(TTU)_{in} + \epsilon_n^* + \varepsilon_{in} \quad (1)$$

Where, $\mathrm{U}_{\mathrm{in}}^{\mathrm{subway}}$ and $\mathrm{U}_{\mathrm{in}}^{\mathrm{bus}}$ are the utilities associated with subway and bus respectively for traveler n and trip i ; $\Delta(\text{travel time})$, $\Delta(\text{TTU})_{\text{in}}$ and $\Delta(\text{travel cost})_{\text{in}}$denote the differences in travel time, travel time unreliability (measured by the standard deviation of travel time) and the travel cost between bus and subway for traveler n and trip i; CBDO and CBDD are two dummy variables that indicate whether the trip has the Central Business District (CBD) as the origination and destination respectively. They also interacted with travel time unreliability (TTU), which is used to estimate the additional effect of travel time unreliability for work-related trips. Finally, we have ϵ_n^*, an error component that is normally distributed and is used to account for the panel nature of the data (Ortúzar and Willumsen, 2024; see also the foundational discussion in Amador et al., 2005), and ε_{in}, the random error term, which is assumed to follow an iid Gumbel distribution leading to a Logit formulation[2]. Therefore, the error component logit estimation is used in this paper.

[2] Given the fact we have panel data, an alternative could have been to implement a model with individual specific error terms to account for the fixed effect for each individual. However, since this would have led to exclusion of travelers who always use

Given the panel nature of our data, an alternative specification would be a model with individual-specific fixed effects. However, since such a model would condition on individual choices, it would exclude travelers who always use the same mode, thereby risking sample selection bias and potentially limiting the generalisability of the results (Ortúzar and Willumsen, 2024). Consequently, the error component (or random effects) logit formulation, which captures unobserved heterogeneity across individuals while retaining the full sample, is adopted as the final and more appropriate model for this study. This choice underscores a critical methodological point: failing to properly account for the panel structure in repeated choice data can lead to biased coefficient estimates and erroneous behavioral inferences, primarily due to the incorrect treatment of error variance across observations (Amador et al., 2005; Ortúzar and Willumsen, 2024).

VOT represents the monetary value that passengers are willing to pay to reduce travel time and VOR indicates the monetary value that passengers are willing to pay to increase reliability, which are the marginal rates of substitution between travel time and travel time reliability and cost, respectively, in the passenger indirect utility function (see Agarwal et al., 2020). The formulae for calculating VOT and VOR respectively are as follows.

$$VOT = \frac{\beta_{tt}}{\beta_{tc}} \tag{2}$$

$$VOR = \frac{\beta_{tr}}{\beta_{tc}} \tag{3}$$

4 Empirical analysis

Error component logit models were estimated for all groups, to analyze differences in travel preferences between peak and off-peak, weekdays and weekends, in addition to exploring the impact of work trips.

4.1 Model results

The differences between bus and subway modes (bus minus subway) were estimated for all passengers, and separately for each of the four groups, using error component logit models. The results are shown in Table 8.

By normalizing the subway utility to zero, we explicitly model bus utility as a deviation from

one travel mode, it posed the risk of introducing bias on the estimation results and hence was not chosen as the final formulation.

this baseline, where the constant term absorbs the systematic preference for subway. The coefficients of travel time, travel time unreliability, and travel cost are negative, and all are significant at the level of 0.01, indicating that longer travel time, higher unreliability of travel time, or increased fares reduce users' willingness to choose to travel by bus. The coefficients of travel time, travel time unreliability, and travel cost are negative, and all are significant at the level of 0.05, indicating that longer travel time, higher unreliability of travel time, or increased fares reduce users' willingness to choose to travel by bus. The absolute value of the cost coefficient for Group 2 is the largest at 4.981, while Group 4 has the smallest at 3.898. This suggests that passengers in Group 2 are most sensitive to fares, while passengers in Group 4 do not care as much.

Meanwhile, The positive CBD coefficients (e.g., CBD_O = 0.701) may reflect unobserved contextual factors, such as the spatial configuration of bus networks in dense business districts offering superior last-meter accessibility compared to subway stations, or targeted policy interventions (e.g., bus priority lanes that are more likely to exist in CBDs) that enhance bus competitiveness for trips originating in the CBD. However, the coefficient of the interaction terms in the models are significantly negative, except for Group 1 where the coefficient of the interaction term with CBD as the origination is positive. The results indicate that for a given mode of transportation, the unreliability of travel time significantly reduces the probability of choosing that mode for work trips relative to other trips.

The inclusion of the panel error term (ϵ_n) in our error component logit model accounts for unobserved individual-specific preferences, such as inherent biases toward subway or bus travel that persist across multiple trips. When this term is added, the estimated coefficients for travel time, cost, and other attributes increase in magnitude because the model more accurately allocates variance.

To validate the behavioral realism of our model, we performed a Monte Carlo simulation. Using the estimated parameters (β_0 = -16.073, σ = 7.212), we generated 1,000,000 draws from the error term distribution, $\epsilon_n \sim N(0, \sigma^2)$. We found that the subset of draws in the upper tail ($\epsilon_n > 15$, 0.11% of the sample) yields a central range of bus choice probabilities (IQR: 6.0%–10.6%) that closely aligns with the empirically observed mode share of 4–9% across traveler clusters, as detailed in Figure 2 and supported by the data in Appendix Table A.1. This quantitative alignment demonstrates that the model successfully replicates the behavior of users with strong unobserved bus preferences.

Group 3 exhibits the largest error term variance ($\sigma = 8.342$), indicating the widest dispersion of latent tastes within this segment. This is consistent with the observed behavior that Group 3 has a larger bus mode share, suggesting a broader potential for behavioral change under service improvements. Conversely, Group 2's more constrained error distribution ($\sigma = 5.966$) reflects greater homogeneity in their unobserved preferences, which aligns with their observed pattern of having the strongest revealed preference for subway (e.g., only 4% bus mode share). These differential patterns substantiate the model's ability to capture authentic behavioral heterogeneity while maintaining theoretical consistency with random utility maximization principles.

The strongly negative constant term ($\beta_0 = -16.073$) reflects the fact that for individuals who use the subway at least once, there is an underlying preference for subway travel in Beijing's transportation system. While this indicates a systematic utility advantage for subway, it does not imply that bus alternatives are behaviorally irrelevant. The model's structure ensures realistic bus choice probabilities (6–10.6%) through the error component (ϵ_n^*), which captures unobserved individual heterogeneity. Even with a negative baseline utility, the right tail of the error distribution generates non-zero bus usage, consistent with observed shares in dataset (5%).

VOT and VOR reflect the trade-off between travel time and variability and transport costs for passengers in their choice of transport mode. Based on equations (2) and (3), we calculated VOT and VOR for each group, as shown in Table 9.

Table 8. Model results

Variables	All passengers	Group 1	Group 2	Group 3	Group 4
$\Delta(travel\ time)$	-0.0890***	-0.0951***	-0.104***	-0.0780***	-0.0803***
	(0.0003)	(0.0006)	(0.0003)	(0.0003)	(0.0002)
$\Delta(travel\ time\ unreliability)$	-0.0725***	-0.06581***	-0.0856***	-0.0681***	-0.0628***
	(0.0004)	(0.0007)	(0.0004)	(0.0004)	(0.0003)
$\Delta(travel\ cost)$	-4.441***	-4.016***	-4.981***	-4.43***	-3.898***
	(0.0116)	(0.0239)	(0.0112)	(0.0115)	(0.0085)
CBD_O	0.701***	0.663***	0.8007***	0.6361***	0.6621***
	(0.0210)	(0.0379)	(0.0191)	(0.0242)	(0.0158)
CBD_D	0.4284***	0.5432***	0.5836***	0.1673***	0.4945***

Variables	All passengers	Group 1	Group 2	Group 3	Group 4
	(0.0240)	(0.0415)	(0.02028)	(0.0242)	(0.0168)
CBD_O×TTU	-0.0289***	0.0101**	-0.0469***	-0.0277***	-0.0200***
	(0.0022)	(0.0037)	(0.00215)	(0.0021)	(0.0016)
CBD_D×TTU	-0.0668***	-0.0396***	-0.0974***	-0.0529***	-0.0524***
	(0.0025)	(0.0042)	(0.00234)	(0.0024)	(0.0018)
constant	-16.073 ***	-13.63***	-16.37***	-18.43***	-13.75***
	(0.7086)	(4.499)	(4.739)	(5.088)	(4.027)
error term	-7.212***	-4.364***	-5.966***	-8.342***	-4.793***
	(0.1937)	（0.018）	（0.057）	（0.2792）	（0.024）
Number of trips	26,791,973	1,159,635	7,156,789	12,180,497	6,295,052
Number of cards	4,892,551	644,174	1,995,370	730,408	1,522,599

Note: ***, **, * represent significant at the significance level of 0.01, 0.05, and 0.1, respectively. Numbers in parentheses indicate standard errors.

Among the four groups of passengers, Group 1 has the highest VOT value of 1.42 yuan/hour (0.20 US$/hour), meaning that passengers in this group are willing to pay 1.42 yuan to shorten their travel time by 1 hour. The VOT for Group 2 is close to Group 4, at 1.25 yuan/hour (0.17 US$/hour). The lowest VOT is for Group 3, well below Group 2 and Group 4, at 1.06 yuan/hour (0.15 US$/hour). These indicate that Group 1 place the most emphasis on travel time, while Group 3 are more sensitive to cost than time and have the lowest VOT. According to the National Bureau of Statistics of China, the average hourly wage in Beijing in 2019 was approximately 23 yuan/hour (3.16 US$/hour). The VOT of Group 2 (1.25 yuan/hour) accounts for approximately 5.4% of the average hourly wage, while Group 3 accounts for approximately 4.6%. Due to low public transportation fares as a result of the Chinese government's financial subsidies, the VOT-to-wage ratios is lower than the ratio of adult transit commuters (15.4% to 15.8%) in the Agarwal et al.'s (2020) study. Because of the economic situation and the public goods nature of the public transport system, the Beijing Municipal Government subsidizes the public transport system to the tune of more than RMB 10 billion a year[3], and actually doesn't rely on the public transport system to make a profit. As a result, Beijing's subway fares are the same as the local price of a bottle of mineral water, and there

3 https://www.beijing.gov.cn/gongkai/caizheng/czzt/2023js/202408/t20240820_3778217.html (in Chinese)

is a certain mental accounting effect (Liu et al., 2021) due to the fact that the fares themselves are so low.

Although the logit model is calculated in minutes, we have converted the values of VOT and VOR to hourly rates as is the norm. Since each hour is 60 minutes, the VOT and VOR can be reasonably converted as shown in Table 9.

Table 9. Estimates of VOT and VOR (yuan/hour)

	All passengers	Group 1	Group 2	Group 3	Group 4
VOT	1.202	1.42	1.25	1.06	1.24
VOR	0.980	0.983	1.03	0.922	0.966
VOR/VOT	0.815	0.692	0.823	0.873	0.781

In terms of VOR, Group 2 places the highest importance on travel time reliability, followed by Group1 and Group 4, with Group 3 placing the lowest. The empirical results presented in Table 9 provide robust evidence for nuanced behavioral differences across passenger groups, reinforcing the credibility of our methodological approach. The observed VOR/VOT ratios—all below 1.0—consistently indicate that travelers generally place a higher monetary value on reducing average travel time than on improving reliability. This pattern aligns with established economic theory and empirical observations from other metropolitan transit systems (Small & Verhoef, 2007; Li, 2017).

Different with the result of VOR, in terms of the relative importance placed on travel time reliability (measured by the VOR/VOT ratio), Group 3 ranks the highest, followed by Group 2 and Group 4, with Group 1 showing the lowest valuation.

In VOR/VOT ratios, Group 3 emerges as a particularly interesting case. Their low absolute valuations for both travel time and reliability may suggest lower average incomes, which would constrain overall willingness-to-pay. However, within this constrained budget, they assign the highest *relative* importance to reliability (VOR/VOT ratio = 0.873). This suggests that for these travelers, securing a predictable journey is a priority when making trade-offs. Their frequent use of CBD and suburban stations (Table 6) is consistent with a 'reliability-seeking' strategy, likely because unpredictable delays pose a disproportionately high risk to their daily schedules, a pattern observed among lower-income, transit-dependent populations who may sacrifice speed for schedule certainty. (Agarwal et al., 2020; Asensio & Matas, 2008; Noland et al., 1998; Noland & Small, 1995).

Group 1's travel behavior is distinctive, defined by their exclusive use of the subway on weekends (0% weekday trips, Table 4) and their travel between tourist hubs and high-speed rail stations (Table 6). This pattern suggests a profile of relatively affluent individuals—such as those who work from home, commute by car, or are tourists—for whom weekend subway travel is a discretionary choice. Their high VOT (1.42 ¥/hour) is consistent with the higher opportunity cost of leisure time for affluent populations. Conversely, their low VOR/VOT ratio (0.692) reflects the inherent flexibility of leisure travel, where the absence of rigid schedule penalties allows for greater tolerance of unreliability. This combination distinguishes them not as typical leisure travelers with low time value, but as a resource-rich segment for whom efficient, high-quality leisure experiences are a priority.

4.2 Variation across time periods

According to the distribution of passenger volume in a day, it is known that 7:30 to 9:00 and 17:30 to 19:00 are peak periods (As shown in Figure 3). Therefore, we divide the two scenarios of peak and off-peak periods to estimate VOT and VOR respectively to explore the variation of passenger preference across time periods. The distribution of trips by group on weekdays and weekends is shown in Table 10. Group 3 travels much more on weekdays than the other groups, and Group 2 and Group 4 travel much more during off-peak periods than during peak periods.

Table 10 Distribution of trips by group on weekdays and weekends

	weekdays		weekends	
	Peak	Off-peak	Peak	Off-peak
All	23.26%	58.72%	2.42%	15.61%
Group 1	0.00%	0.00%	0.43%	3.46%
Group 2	7.31%	19.24%	0.00%	0.00%
Group 3	12.58%	26.38%	1.08%	5.82%
Group 4	3.37%	13.10%	0.91%	6.32%

As shown in Table 11, except for Group 2, the value of travel time is greater in the peak period than in the off-peak period. The value for time reliability is smaller for all groups in the peak period than in the off-peak period. This may be due to the increased uncertainty of travel time caused by peak-hour congestion during commuting. Therefore, passengers who travel during peak periods may have more rigid schedules and place a relatively higher value on travel time. Those who already choose to stagger their trips generally have a higher VOT during the peak hours, but with different degrees of importance for each group.

Table 11. Estimates of VOT and VOR during peak and off-peak periods (Yuan/Hour)

	All passengers		Group 1		Group 2		Group 3		Group 4	
	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak
VOT	1.06	1.21	1.0291	1.49	1.31	1.23	0.856	0.965	1.109	1.269
VOR	0.651	1.18	0.5382	1.20	0.651	1.25	0.696	1.23	0.701	1.276
VOR/VOT	0.616	0.972	0.5230	0.804	0.498	1.01	0.813	1.28	0.632	1.001

4.3 Difference between weekdays and weekends

On the basis of different time periods, VOT and VOR are estimated separately for weekday and weekend scenarios. The travel preferences of passengers are consistent for all groups.

Weekday travel patterns reveal a consistent trend across all groups with available data (Table 12): the Value of Travel Time (VOT) is generally higher during off-peak periods than during peak hours. This may reflect that travelers in less congested periods place a greater marginal value on saving time for discretionary activities, once freed from the rigid constraints of rush-hour commutes. In contrast, the Value of Reliability (VOR) presents a more nuanced pattern. While remaining elevated during peak periods, it increases markedly during off-peak hours for Groups 2, 3, and 4. Notably, Group 3 exhibits the highest VOR/VOT ratio (1.398) in off-peak periods, indicating that these passengers attach greater relative importance to reducing travel time variability than to saving average travel time. This finding is consistent with the overall model results (Table 9), where Group 3 also shows the highest VOR/VOT ratio (0.873), underscoring that a strong preference for reliability is a robust and defining characteristic of this segment.

Weekend travel patterns further highlight behavioral segmentation among the groups (Table 13). Group 2 shows no subway usage on weekends, reinforcing their identity as weekday-oriented travelers, likely bound by work-related schedules. Conversely, Group 1 consists exclusively of weekend users, with travel concentrated during leisure-oriented time windows; however, their off-peak estimates are statistically unreliable due to data limitations, as noted in the footnote 4, and are thus excluded from this behavioral interpretation. Between these two extremes, Groups 3 and 4 demonstrate continued subway usage across both weekdays and weekends, suggesting a broader dependence on public transport for diverse travel needs. These presence-based patterns offer a clear behavioral segmentation: from weekday-constrained (Group 2) and leisure-focused (Group 1) users to cross-day reliant passengers (Groups 3 and 4), illustrating how travel purpose and system dependency shape valuation structures.

Across all analyses, the VOR/VOT ratio serves as a key metric for understanding the relative importance of reliability in passengers' decision-making. Its consistency across modeling approaches reinforces that travel behavior is not merely period-specific but reflects underlying and stable passenger preferences.

Table 12. Estimates of VOT and average VOR during peak and off-peak periods on weekdays (Yuan/Hour)

	All passengers		Group 1		Group 2		Group 3		Group 4	
	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak
VOT	1.0577	1.1529	-	-	1.1784	1.2341	0.8554	0.9648	1.1105	1.1877
VOR	0.8652	0.8314	-	-	0.8741	1.3778	0.7813	1.3488	0.7908	1.3655
VOR/VOT	0.8197	0.7211	-	-	0.7418	1.1165	0.9134	1.3980	0.7122	1.1497

Note: Group 1 has no trip record on weekdays.

Table 13. Estimates of VOT and average VOR during peak and off-peak periods on weekends (Yuan/Hour)

	All passengers		Group 1		Group 2		Group 3		Group 4	
	Peak	Off-peak	Peak	Off-peak[4]	Peak	Off-peak	Peak	Off-peak	Peak	Off-peak
VOT	1.1422	1.3589	1.3960	0.0275	-	-	0.9601	1.1311	1.1826	1.4210
VOR	0.4304	0.9926	0.4916	0.0188	-	-	0.4303	0.9349	0.4216	1.0198
VOR/VOT	0.3768	0.7304	0.3522	0.6830	-	-	0.4482	0.8265	0.3565	0.7176

[4] The estimates for Group 1's weekend off-peak period are not statistically reliable due to the conceptual inappropriateness of the peak/off-peak classification for leisure travel and the limited data in this segment. These values are therefore not recommended for behavioral interpretation.

Note: Group 2 has no trip record on weekends.

4.4 Differences for work trips

Passengers' preferences for transport service attributes are often related to the purpose of the trip. Commuters may face greater "penalties" for travel delays due to work schedule constraints. Travel time unreliability refers to the fact that passengers cannot accurately predict how long a trip will last. A more unreliable travel time means a higher probability that the actual travel time does not match the estimated travel time due to random factors.

Here, we assume that trips that start or end in the CBD are work-related. Table 14 shows the estimated VOT and VOR for passengers departing from or arriving at the CBD. The first column indicates the results for all trips into or out of the CBD, while the second and third columns indicate the results for departures from and arrivals at the CBD, respectively. The VOR of passengers entering the CBD is high, with a value of 1.764 for VOR/VOT, indicating that passengers place much more importance on reliability than on travel time when using the CBD as their destination. The VOT value for passengers leaving the CBD is slightly higher, with a value of 0.731 for VOR/VOT. Most trips from the CBD area are home from work, when there are no more time constraints, so passengers are less sensitive to travel time reliability. However, after a busy day, the desire to go home early to rest has grown, so the willingness to pay for a shorter travel time has increased. These results suggest that the purpose of the trip affects passengers' preferences for service attributes. For CBD-related trips (both entering and leaving), passengers' VOR was higher than the VOR for all trips (Table 9), indicating that passengers are very sensitive to travel time reliability for work trips, consistent with the results of the study by Agarwal et al. (2020) on commuters' service preferences.

Table 14. VOT and VOR for passengers on trips related to the CBD area (Yuan/Hour)

	All passengers	Entering the CBD	Leaving the CBD
VOT	1.202	1.098	0.843
VOR	1.295	1.937	0.617
VOR/VOT	1.077	1.764	0.731

5 Discussion

5.1 Analysis of the results

This paper explores the general travel characteristics of different passenger groups using public transportation smart card transaction data and estimates each group's preferences for travel time, travel time reliability, and travel cost. The dataset contains all trip record data from the Beijing public transportation system in November 2019, and more than 66 million trip records are retained through data cleaning and preliminary processing.

First, the feature variables were set from three aspects of travel frequency, temporal characteristics and spatial characteristics, and the k-means++ clustering algorithm was applied to classify the crowd into four groups. Analysis of the travel characteristics of each group of passengers reveals that the travel characteristics of Group 3 closely match those of commuters. The departure time of their first trip of a day almost always falls during the morning peak, and the departure time of their last trip likewise almost coincides with the evening peak time period. Furthermore, they take the subway with high frequency and are thus the largest target group for subway peak hour congestion management, and they completed the highest percentage of trips at 47.32%. The most frequent destination for Group 3 is the CBD area, accounting for approximately 49.75% of the trips to the CBD area. Passengers in Group 1 usually travel by subway only on non-working days, and the destinations of their trips are mostly shopping areas and tourist attractions. In terms of the most frequent departure places, subway stations near colleges and universities rank high, indicating that the travel characteristics of Group 1 are consistent with the behavioral characteristics of most college students. For Group 2 and Group 4, the characteristics of trips made by subway on weekdays are relatively similar, with most passengers taking the subway only once in a day for a longer trip, but less frequently. These two groups have more trips that start or end at high-speed rail stations, possibly including many passengers who need to make cross-province trips for reasons such as business or tourism. The peak period of trips in a day for Group 2 and Group 4 is concentrated around 18:00, which has a high overlap with the evening peak time period. Among the trips starting from the CBD, Group 2 and Group 4 account for 28.24% and 20.31% respectively, suggesting that these two groups include some commuters who take other alternative transportation modes to work and the subway to return home from work. This research found that variations in travel characteristics across groups are significant, and one-size-fits-all incentives may not be able to cope with these differences, in addition to previous research (Halvorsen et al., 2016; Henn et al., 2011; Ma & Koutsopoulos, 2019) suggesting that response rates to one-size-fits-all measures are relatively low. Classifying passengers by their travel characteristics followed by group analysis is potentially more conducive to improving the effectiveness of targeted interventions aimed at changing their travel behavior.

Then, we estimated the public transport mode choice models for each group. The odds ratios of travel time, travel time unreliability, and travel cost are all less than 1 and statistically significant, indicating that longer and more variable travel time as well as higher fares will reduce the possibility of public transport users choosing to take the bus.

Based on the results of the error component logit model, we calculated the VOT and VOR for each group of passengers. There are noticeable differences in the value of travel time and reliability for each group. Among the four groups of passengers, Group 1 has the highest VOT (1.42 ¥/h) and Group 4 the second (1.24 ¥/h), implying that they place a high value on travel time. For Group 1 (primarily leisure or tourist trips), this may reflect the high opportunity cost of their limited discretionary time, motivating them to minimize travel duration to maximize recreational activities. Group 3 exhibits the lowest VOT and VOR but the highest VOR/VOT ratio, underscoring their distinct preference for travel time reliability over speed, a pattern consistent with their behavioral profile discussed earlier. Group 1 attaches the least emphasis to relative importance of travel time reliability (VOR/VOT = 0.692), reflecting their flexible leisure-oriented travel patterns.

Estimating travel preferences at different time periods reveals that passengers travelling during peak periods have a higher VOT, perhaps because they have a more rigid schedule. Whereas passengers who choose to stagger their trips generally have a higher VOR, they want a more certain trip. The difference in travel preferences between weekdays and weekends suggests that passengers are more likely to stagger their trips when they have more flexible schedules.

By analyzing work trips, we found that trip purpose influences passengers' preferences regarding service attributes. The unreliability of travel time reduces the likelihood of choosing a certain travel mode for work trips compared to other trips. Group 3 has the highest percentage of trips related to CBD (starting from or arriving at CBD) at 49.21%, while Group 1 has the lowest percentage at 2.85%.The VOR/VOT ratio for trips entering the CBD is 1.764 (Table 14), significantly higher than for trips leaving the CBD (0.731). This suggests that the constraints of commuting time may lead to commuters facing a greater cost of being late, so the sensitivity to reliability is much greater than that of travel time. In contrast, for trips departing the CBD, VOR drops to 0.617 while VOT rises to 0.843 (Table 14), with a VOR/VOT ratio of 0.731. Travelers departing from the CBD are likely to be commuters returning home from work. At this point, they are less sensitive to travel time reliability and more attentive to travel time because there are no more time constraints and an increased desire for an early return home. Overall, passengers place a higher value on reliability for work trips than for other trips, whether they begin or end in the CBD.

5.2 Policy implication

Overcrowding on subway systems during peak periods is a challenge for public transport management in numerous cities. Many studies have demonstrated that the effectiveness of "one-size-fits-all" demand management measures can be very poor (Halvorsen et al., 2016; Henn et al., 2010, 2011; Ma et al., 2020; Wang et al., 2020). Interventions that focus on "cutting the peak and filling the trough", that is, shifting subway passengers who travel during peak periods to off-peak periods, may have more impact on passengers closer to off-peak times (when the time required for shifting is short), but as the shifting time becomes longer, the response of passengers to the incentive decreases sharply (Halvorsen et al., 2020; Henn et al., 2011; Peer et al., 2016; Zou et al., 2019). Advances in information and communications technology have greatly expanded the responsiveness of transportation systems to user needs and the rise of the modern sharing economy, making MaaS (Mobility-as-a-Service) embrace disruptive social trends that could dramatically change the mode of operation and type of service of public transportation (Hensher, 2017; Mulley & Kronsell, 2018; Smith et al., 2022). Demand responsive transport (DRT) shows great potential for improving the operational efficiency and service level of public transportation systems. It aims to integrate users with similar travel needs and provide point-to-point, customized bus services through route planning and vehicle dispatching technologies, thereby improving service quality. Compared with traditional fixed-route buses, DRT is a good way to improve service reliability and reduce waiting time for users, but the high fare is the main obstacle to increase patronage (Alonso-González et al., 2018; Coutinho et al., 2020; Currie & Fournier, 2020; Narayan et al., 2020; Perera et al., 2020; Ryley et al., 2014).

For passengers who place less importance on service reliability, the main incentives include monetary rewards, accumulating points for prizes, other service rewards (e.g., dining coupons, etc.), and providing congestion information, with Wang et al. (2018) showing that monetary rewards are the most effective. In addition, Huang et al. (2016) proposed a nonlinear OD-based bus fare structure in which the fare is determined based on the Euclidean distance function between the origin and the destination, regardless of the number of interchanges. The results of the study show that this fare structure can lead to a more even distribution of passengers on each bus route and increase the number of passengers who choose to transfer routes. At the same time, the OD-based fare structure reduces the monetary cost for passengers riding the interchange routes which can attract a portion of price-sensitive passengers.

Previous subway congestion studies have focused more on the trade-off between passenger travel time and fare, resulting in a situation where it is usually the price-sensitive passengers who respond to incentives, while the monetary cost of making the shift for the less price-sensitive passengers is unaffordable for transit operators. Therefore, other solutions need to be found. Our findings show that passengers with low price sensitivity can be divided into two main segments, one for leisure and entertainment purposes and the other for commuting. They are less sensitive to pre-peak incentives and discounts, but shorter travel time and improved reliability may have a better effect on changing their travel behavior. For example, running DRT or designating bus lanes during peak periods can both reduce travel time and improve the reliability of bus service, thereby attracting more commuters to switch to bus travel and reducing congestion on the subway system.

Recognizing the demand patterns and passengers' preferences associated with congestion can inform the development of intervention policies and help design policies for higher response rates and improved service quality. The heterogeneity of users' travel patterns and preferences suggests that different types of passengers require targeted incentive programs. "One-size-fits-all" transportation policies may not achieve effective congestion management, so it is worth considering combining multiple incentive schemes and transportation modes to design more personalized travel services. The integration of smartphones and automated fare collection systems provides support for the deployment of personalized transportation strategies. According to the results of this paper and referring to previous practical experiences on PTDM, we propose the following incentives for each type of group, in order to provide a reference for alleviating subway congestion during peak hours and improving the overall efficiency of the public transportation system.

Passengers in Group 1 have high willingness to pay for shortening travel time and low sensitivity to fares, so they prefer to travel by subway and have a low response rate to monetary incentives. However, in terms of travel characteristics, almost all of their trips occur on non-working days, while subway congestion usually occurs on working days, so the impact of their trips on subway peak congestion is minimal and there is no need to try to change their travel behavior.

The relative high price sensitivity of Group 2 suggests that monetary incentives are attractive to them. However, their high VOT (1.25 ¥/h) means that travel time is the main resistance to change their travel patterns. The possible reason for this is that their trips are generally relatively long, leading to a higher willingness to pay for shorter travel time.

One solution to incentivize this group is to make staggered trips through off-peak subway fare discounts. The advantage of this option is that it guarantees shorter travel time, while taking advantage of their higher monetary sensitivity to motivate them to travel during off-peak periods. The disadvantage, however, is that is that as the time required for peak avoidance increases, the response rate of passengers decreases. In addition, passengers who would otherwise travel during off-peak periods are rewarded for not contributing to congestion management, leading to low cost-benefit efficiency.

Another solution is to induce peak travelers to change their travel behavior by shifting to off-peak periods to obtain points rewards that can be redeemed for gifts or lottery tickets. This option will improve cost-benefit efficiency, but the reward form is not as intuitive as monetary rewards and requires more detailed design and promotion of the point redemption process and rules. We also note that the ODs of this group are more fixed and concentrated, meaning that most passengers usually travel between a small number of ODs. This suggested that route-based incentives could further optimize network load distribution.

Group 3, the key commuter group, shows the lowest absolute VOT and VOR but the highest VOR/VOT ratio. This suggests that while their overall willingness-to-pay is constrained, reliability is their paramount concern. Thus, service improvements targeting predictability are crucial for managing their peak-hour travel.

The development of demand-responsive buses may be effective in attracting them to switch their travel mode, especially since 49.21% of their trips begin or end in the CBD (Table 14). Meanwhile, it has been demonstrated that there is a higher success rate of many-to-few operating models (multiple locations at the end of a trip, but small enough to be effectively managed) compared to more comprehensive and complex many-to-many service types (where passengers can come from or go to any location) (Currie & Fournier, 2020). Therefore, developing a DRT system with an end location in the CBD can further improve the feasibility of the program. In addition, designating bus lanes during peak periods is another option to attract passengers to transfer to buses. Running demand-responsive buses requires the development of new response systems and applications, as well as the scheduling of dispatchable vehicles, which is more costly but provides a higher level of service. Designated bus lanes do not need to change the existing bus routes and is less costly. Correspondingly, the level of service is lower than that of DRT.

Passengers in Group 4 have a relatively high sensitivity to fares, so monetary incentives can change their travel behavior to some extent. Their moderate VOT (1.24 ¥/h) and VOR (0.966 ¥/h) suggest balanced preferences for time and reliability. Also, they travel at a higher frequency both on weekdays and weekends. Accordingly, they can be encouraged to switch to buses during peak hours or take the subway during off-peak hours through monetary incentives. For example, a monthly bus card discount might have the potential to attract them to travel by bus, thus reducing subway congestion and increasing the utilization of buses. Optimization of the bus fare structure and measures such as off-peak subway fare discounts or lotteries can also be chosen depending on the operator’s considerations.

The systematic variations in VOT and VOR across passenger groups and time periods collectively point to rigid schedule constraints as a key driver of peak-hour congestion. This is exemplified by Group 3, the core commuter segment, whose high VOR/VOT ratio signifies a substantial cost associated with schedule delays. The marked shift from concentrated weekday peak travel to dispersed weekend patterns further indicates that flexibility, when allowed, naturally leads to trip staggering. Consequently, alongside targeted incentives, promoting flexible working hours emerges as a foundational strategy to address the root cause of congestion, potentially shifting travel demand across segments.

6 Conclusion

This paper explores the heterogeneity of travel patterns and travel preferences of Beijing subway users and proposes targeted incentive schemes. Using smart card data, we construct clustering indicators in terms of travel frequency, temporal characteristics, and spatial characteristics to identify the travel patterns of different passenger groups, which is not possible with traditional survey data. For the measurements of VOT and VOR, the use of revealed preference data improves the accuracy and validity of the results, extending insight into the development and evaluation of the effectiveness of congestion management measures in the Beijing subway. The analysis of passenger groups with different travel purposes provides a better understanding of the full picture of peak-hour travelers. The significant differences in travel characteristics and travel preferences of passengers suggest that future TDM should focus on passenger-centric approaches, and a combination of diverse transportation modes and incentives will be more effective in improving the efficiency of the transportation system and passenger satisfaction.

Due to privacy protection, personal information is not available in the smart card dataset. Further research is needed on the travel characteristics and preferences of public transportation users. Combining smart card data with SP survey data can be useful for gaining a more definitive understanding of users' behavioral purposes and allows for better control of other factors such as lifestyle characteristics and sociodemographic variables in the future study (Wang et al. 2025). In addition, further segmentation of passengers by adding more sources of data, such as weather data and carriage crowding information, is also a potential direction for future research. In the future, with the information on VOT and VOR of passengers, transportation managers can design targeted incentives for different groups, and it is worthwhile analyzing the behavior change of passengers and compare the effect with the "one-size-fits-all" policy.

This study incorporates smart card data but without experimental setting, thus this study avoids the bias of SP approaches, but the data collection environment is not fully controllable, therefore it has its own drawbacks as well. Future study approach with proper experimental scenario setting could capture more details in the behavioural response.

The constraints on data accessibility and server processing capacity precluded the utilisation of more than one month's data for processing in the present study. Consequently, the medium- to long-term travel characterisation, such as seasonal impacts, could not be conducted. It is recommended that future studies adopt a longer-term perspective and incorporate a greater number of factors in their analysis.

While the direct results of this paper apply to Beijing, which is considered a microcosm of other large cities in China and the rest of the world (Wang et al., 2020), the framework to use smart card data for capturing the heterogeneity in VOT and VOR among different groups of passengers can be used for subway congestion management in other large cities.

Acknowledgments

Professor Yacan Wang was supported by the Natural Science Foundation of China [grant number 72474018]. Professor Huiyu Zhou was supported by National Social Science Foundation of China [grant number 22BJY211]. Professor Charisma Choudhury and Dr Thomas Hancock's time was supported by the UKRI Future Leader Fellowship [MRT020423/1]. We thank Beijing Mass Transit Railway Operation Corporation Limited for providing the anonymous data used for this research.

The authors would also like to express their sincere gratitude to the anonymous reviewers for their constructive and insightful comments, which have greatly improved the quality of this manuscript. We are particularly indebted to the reviewer who provided exceptionally thorough and valuable suggestions, especially regarding the rigorous assessment of cluster stability, which strengthened our methodological robustness, and for the guidance on more closely linking methodological insights with practical policy implications. Their detailed engagement has been instrumental in shaping the final version of this paper.

References

Agarwal, S., Diao, M., Keppo, J., & Sing, T. F. (2020). Preferences of public transit commuters: Evidence from smart card data in Singapore. *Journal of Urban Economics*, *120*, 103288. https://doi.org/10.1016/j.jue.2020.103288

Ali, A., Kim, J., & Lee, S. (2015). Travel behavior analysis using smart card data. *KSCE Journal of Civil Engineering*, *20*. https://doi.org/10.1007/s12205-015-1694-0

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. https://doi.org/10.1016/j.trc.2017.12.016

Amador, F. J., González, R. M., & Ortúzar, J. de D. (2005). Preference Heterogeneity and Willingness to Pay for Travel Time Savings. Transportation, 32(6), 627–647. https://doi.org/10.1007/s11116-005-3734-y

Asensio, J., & Matas, A. (2008). Commuters' valuation of travel time variability. *Transportation Research Part E: Logistics and Transportation Review*, *44*(6), 1074–1085. https://doi.org/10.1016/j.tre.2007.12.002

Bagchi, M., & White, P. R. (2005). The potential of public transport smart card data. *Transport Policy*, *12*(5), 464–474. https://doi.org/10.1016/j.tranpol.2005.06.008

BART. (2019). *BART Perks Phase II Evaluation Report*. San Francisco Bay Area Rapid Transit District.https://www.bart.gov/sites/default/files/docs/Perks%20Phase%20II%20-%20FTA%20Final%20Report.pdf

Bates, J., Polak, J., Jones, P., & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2), 191–229. https://doi.org/10.1016/S1366-5545(00)00011-9

Bhat, C. R., & Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B: Methodological*, *40*(9), 709–730. https://doi.org/10.1016/j.trb.2005.09.008

Carrion, C., & Levinson, D. (2012). Value of travel time reliability: A review of current evidence. *Transportation Research Part A: Policy and Practice*, *46*(4), 720–741. https://doi.org/10.1016/j.tra.2012.01.003

Currie, G. (2009). Exploring the impact of the "Free Before 7" campaign on reducing overcrowding

Dixit, V. V., Ortmann, A., Rutström, E. E., & Ukkusuri, S. V. (2017). Experimental Economics and choice in transportation: Incentives and context. *Transportation Research Part C: Emerging Technologies*, *77*, 161–184. https://doi.org/10.1016/j.trc.2017.01.011

Fifer, S., Rose, J., & Greaves, S. (2014). Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it? Transportation Research Part A: Policy and Practice, 61, 164–177. https://doi.org/10.1016/j.tra.2013.12.010

Gärling, T., & Schuitema, G. (2007). Travel Demand Management Targeting Reduced Private Car Use: Effectiveness, Public Acceptability and Political Feasibility. *Journal of Social Issues*, *63*(1), 139–153. https://doi.org/10.1111/j.1540-4560.2007.00500.x

Goulet-Langlois, G., Koutsopoulos, H. N., and Zhao, J. (2016). Inferring patterns in the multi-week activity sequences of public transport users. Transportation Research Part C: Emerging Technologies. https://doi.org/64:1-16. 10.1016/j.trc.2015.12.012

Greene-Roesel, R., Castiglione, J., Guiriba, C., & Bradley, M. (2018). BART Perks: Using Incentives to Manage Transit Demand. *Transportation Research Record: Journal of the Transportation Research Board*. https://doi.org/10.1177/0361198118792765

Halvorsen, A., Koutsopoulos, H. N., Lau, S., Au, T., & Zhao, J. (2016). Reducing Subway Crowding: Analysis of an Off-Peak Discount Experiment in Hong Kong. *Transportation Research Record: Journal of the Transportation Research Board*, *2544*(1), 38–46. https://doi.org/10.3141/2544-05

Halvorsen, A., Koutsopoulos, H. N., Ma, Z., & Zhao, J. (2020). Demand management of congested public transport systems: A conceptual framework and application using smart card data.

Transportation, *47*(5), 2337–2365. https://doi.org/10.1007/s11116-019-10017-7

Henn, L., Douglas, N., & Sloan, K. (2011). Surveying Sydney rail commuters' willingness to change travel time. *World Transit Research*, 16.

Henn, L., Karpouzis, G., & Sloan, K. (2010). A review of policy and economic instruments for peak demand management in commuter rail. *World Transit Research*, 21.

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, *52*(1), 258–271. https://doi.org/10.1016/j.csda.2006.11.025

Hollander, Y. (2006). Direct versus indirect models for the effects of unreliability. *Transportation Research Part A: Policy and Practice*, *40*(9), 699–711. https://doi.org/10.1016/j.tra.2005.12.007

Huang, D., Liu, Z., Liu, P., & Chen, J. (2016). Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. *Transportation Research Part E: Logistics and Transportation Review*, *96*, 1–19. https://doi.org/10.1016/j.tre.2016.10.004

Kusakabe, T., & Asakura, Y. (2014). Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies*, *46*, 179–191. https://doi.org/10.1016/j.trc.2014.05.012

Lam, T. C., & Small, K. A. (2001). The value of time and reliability: Measurement from a value pricing experiment. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2), 231–251. https://doi.org/10.1016/S1366-5545(00)00016-8

Li, S.-M., & Wong, F. C. L. (1994). The effectiveness of differential pricing on route choice. *Transportation*, *21*(3), 307–324. https://doi.org/10.1007/BF01099216

Li, Z., 2017. Uncertainty aversion and its role in travel decision making under uncertainty. *International Journal of Strategic Decision Sciences,*8(1), 1-12. https://doi.org/10.4018/IJSDS.2017010101

Li, W., Yan, X., Li, X., & Yang, J. (2020). Estimate Passengers' Walking and Waiting Time in Metro Station Using Smart Card Data (SCD). *IEEE Access*, *8*, 11074–11083. https://doi.org/10.1109/ACCESS.2020.2965155

Liu, H. X., He, X., & Recker, W. (2007). Estimation of the time-dependency of values of travel time and its reliability from loop detector data. *Transportation Research Part B: Methodological*, *41*(4), 448–461. https://doi.org/10.1016/j.trb.2006.07.002

Liu, H. X., Recker, W., & Chen, A. (2004). Small. *Transportation Research Part A: Policy and Practice*, *38*(6), 435–453. https://doi.org/10.1016/j.tra.2004.03.003

Liu, S. , Zhu, J. , Easa, S. , & Xu, Y. . (2021). Travel choice behavior model based on mental accounting of travel time and cost. *Journal of advanced transportation*. https://doi.org/10.1155/2021/4219396.

Ma, Z., & Koutsopoulos, H. N. (2019). Optimal design of promotion based demand management strategies in urban rail systems. *Transportation Research Part C: Emerging Technologies*, *109*, 155–173. https://doi.org/10.1016/j.trc.2019.10.008

Ma, Z., Koutsopoulos, H. N., Liu, T., & Basu, A. A. (2020). Behavioral response to promotion-based public transport demand management: Longitudinal analysis and implications for optimal promotion design. *Transportation Research Part A: Policy and Practice*, *141*, 356–372. https://doi.org/10.1016/j.tra.2020.09.027

Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. *Journal of Physics: Conference Series*, *1361*(1), 012015. https://doi.org/10.1088/1742-6596/1361/1/012015

Noland, R. B., & Small, K. A. (1995). Travel-time uncertainty, departure time choice, and the cost of morning commutes. *Transportation Research Record*, *1493*, 150–158.

Noland, R. B., Small, K. A., Koskenoja, P. M., & Chu, X. (1998). Simulating travel reliability. *Regional Science and Urban Economics*, *28*(5), 535–564. https://doi.org/10.1016/S0166-0462(98)00009-X

Ortúzar, J. de D., & Willumsen, L. G. (2024). Modelling Transport (5th ed.). John Wiley & Sons.

Peer, S., Knockaert, J., & Verhoef, E. T. (2016). Train commuters' scheduling preferences: Evidence from a large-scale peak avoidance experiment. *Transportation Research Part B: Methodological*, *83*, 314–333. https://doi.org/10.1016/j.trb.2015.11.017

Pluntke, C., & Prabhakar, B. (2013). INSINC: A Platform for Managing Peak Demand in Public Transit. *SUSTAINABLE URBAN TRANSPORT*, 9.

Small, K. A., Winston, C., & Yan, J. (2005). Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. *Econometrica*, *73*(4), 1367–1382.

Small, K. A., & Verhoef, E. T. (2007). The economics of urban transportation. Routledge. https://doi.org/10.4324/9780203642306

Tillema, T., Ben-Elia, E., Ettema, D., & van Delden, J. (2013). Charging versus rewarding: A comparison of road-pricing and rewarding peak avoidance in the Netherlands. *Transport Policy*, *26*, 4–14. https://doi.org/10.1016/j.tranpol.2012.01.003

Tseng, Y.-Y., & Verhoef, E. T. (2008). Value of time by time of day: A stated-preference study. *Transportation Research Part B: Methodological*, *42*(7), 607–618. https://doi.org/10.1016/j.trb.2007.12.001

Van Loon, R., Rietveld, P., & Brons, M. (2011). Travel-time reliability impacts on railway passenger demand: A revealed preference analysis. *Journal of Transport Geography*, *19*(4), 917–925. https://doi.org/10.1016/j.jtrangeo.2010.11.009

Wang, Y., Ettema, D., Zhou, H., & Sun, X. (2018). Understanding peak avoidance commuting by subway: An empirical study in Beijing. *International Journal of Logistics Research and Applications*, *21*(6), 597–613. https://doi.org/10.1080/13675567.2018.1466871

Wang, Y., Wang, Y., & Choudhury, C. (2020). Modelling heterogeneity in behavioral response to peak-avoidance policy utilizing naturalistic data of Beijing subway travelers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *73*, 92–106. https://doi.org/10.1016/j.trf.2020.06.016

Wang, Y., Hancock, T.O., Wang, Y. and Choudhury, C., 2025. Modelling residential relocation behaviour combining passive revealed preference data and stated preference survey data. Transport Policy, p.103789. https://doi.org/10.1016/j.tranpol.2025.103789

Zhu, S., Jiang, G., & Lo, H. K. (2017). Capturing Value of Reliability through Road Pricing in Congested Traffic under Uncertainty. *Transportation Research Procedia*, *23*, 664–678. https://doi.org/10.1016/j.trpro.2017.05.037

Zou, Q., Yao, X., Zhao, P., Wang, Z., & Yang, T. (2019). Measuring Retiming Responses of Passengers to a Prepeak Discount Fare by Tracing Smart Card Data: A Practical Experiment in the Beijing Subway. *Journal of Advanced Transportation*, *2019*, e6873912.

https://doi.org/10.1155/2019/6873912

Appendix

Table A.1 Composition of Analyzed and Excluded Travelers

	User type	Number of trips	Number of cards	Card percentage in sub types
Subway Trip Cards	Multimodal candidates	25.3 million	4.89 million	61.7%
	Unmatched in OD (match step 2)	40.7million	3.03 million	38.2%
	Subtotal	**66 million**	**7.92 million**	---
Bus Trip Cards	Multimodal	1.4 million	0.96 million	5.8%
	Unmatched in OD (match step 2)	53.4 million	10 million	60.5%
	Bus-only (Removed in match step 3)	30.2 million	5.56 million	33.7%
	Subtotal	**85 million**	**16.52 million**	---
Remained Cards/Trips	Deduplicated	26.79 million	4.89 million	---

Note: *Multimodal candidate users (4.89M) contributed all 1.4M matched bus trips, averaging 0.32 bus trips per card. The 30.2M excluded bus trips belong to Bus-Only users (5.56M cards) who never used subway, thus irrelevant to our subway users' mode choice analysis.*

The data in Table A.1 identifies 4.89 million subway users who traveled in OD pairs with both subway and bus options (making them *OD-matched multimodal candidates*). Of these, only 0.96 million cards (19.6%) were observed to actually used bus services (generating 1.4 million bus trips). Therefore, there are 3.93 million cards (80.4%) that used subway exclusively despite available bus alternatives. This substantial retention of subway usage among users with competing mode options highlights the need to separate *infrastructure-enabled multimodality* from *actual mode-shifting behavior* in transit studies. Deduplicated cards in the table means we counted each unique anonymous user ID only once for remaining cards. This was necessary because our trip-based data contained multiple records per user. Counting unique users, as opposed to total trips, is essential for accurate population-level analysis (e.g., determining what percentage of users are multimodal).