Crowding

Yizhen Gu Qu Tang Kai Wu Ben Zou*

December 2025

(Click HERE for the most recent draft)

Abstract

Crowding is a common disamenity that exists in many settings, but little is known about its monetary cost. This paper estimates the willingness-to-pay (WTP) to avoid crowding in public transportation with a revealed preference framework. We leverage an off-peak pricing discontinuity in the Beijing Subway that generates exogenous temporal variation in price. Passengers traveling between a pair of stations choose the optimal departure time, trading off between the price, the expected level of crowding, and the deviation from the ideal time of travel. We devise a novel approach to allocate passengers to trains and calculate real-time crowding. Crowding is correlated with unobserved demand shocks that make traveling in a station pair at a specific time particularly appealing. To address the endogeneity in crowding, we construct an instrumental variable based on the number of overlapping trips that start from and end in different stations, and are thus driven by plausibly unrelated demand shocks. We estimate the marginal WTP to reduce in-train crowding by one passenger per square meter to be 5 cents RMB per minute, or about 2 RMB for a typical 40-minute subway ride. With an average crowding of 3.4 persons per square meter, the trip generates a crowding externality of around 7 RMB. High-income passengers are less price sensitive but have a higher WTP to avoid crowding. Using these estimates, we evaluate the welfare impacts of alternative policies. An optimal crowding tax of 5.5 RMB raises welfare but reduces ridership and disproportionately harms low-income passengers. In contrast, a two-class configuration improves welfare for both high- and low-income passengers by soliciting self-selection. Quantity control performs much worse in passenger welfare and system revenue, and generates large externality on surface roads. Our results underscore the large

*Gu: Peking University, yizhengu@phbs.pku.edu.cn; Tang: Jinan University, qutang@jnu.edu.cn; Wu: University of Texas at Austin, kaiwu@utexas.edu; Zou: Purdue University, zou136@purdue.edu. We thank Marika Cabral, Yue Cao, Gilles Duranton, Matthew Freedman, Ryuya Ko, Shanjun Li, Hayden Parsley, Cody Tuttle, Qinshu Xue, and workshop and conference participants at the Chinese University of Hong Kong (Shenzhen), Shanghai University of Finance and Economics, the University of Texas, Urban Economics Association Annual Meeting for comments. Yunlei Tang, Danhe Tang, and Xiangjie Zhu provided excellent research assistance. Gu gratefully acknowledges funding from the Natural Science Foundation of China (Grant No. 72273008). All remaining errors are our own.

welfare costs of subway crowding and highlight differentiated pricing as a more efficient and equitable policy tool.

Keywords: Crowding externality, Public transit, Travel demand.
JEL Codes: R41, R48, L92, D62

1 Introduction

Crowding, defined as a large number of people crammed in a small space, is a common experience in daily life. People constantly complain about crowded venues such as restaurants, stadiums, classrooms, and public transportation, as crowding brings discomfort. However, it is unclear how much people are willing to pay to avoid crowding.

There are at least two empirical challenges to credibly estimating the disutility from crowding. First, crowding is endogenous. It is formed when many people respond to the same, often unobserved, demand shock and convene to the same location. Second, crowding is correlated with other desired or undesired features that are unobservable to the researcher. For example, a crowded restaurant may have slower service; buses are more likely to be crowded when roads are congested, leading to delays in travel. Third, observed variations in crowding are often an endogenous outcome of differentiated offerings of price and features. In those settings, less crowded options are offered for higher prices, such as business classes on airplanes and skyboxes in stadiums, those options typically also come with other perks, such as better services or a symbol of status. In some contexts, crowding may be perceived as an amenity, as with the "vibes" at busy bars and live music concerts. Existing studies predominantly rely on surveys and use stated preferences to measure disutility from crowding. However, what people *state* may differ from how they actually *behave*.

This paper estimates the marginal willingness to pay (MWTP) to avoid crowding in the public transit, focusing on the Beijing Subway. Crowding is an important supply-side challenge in many public transit systems, and quantifying passengers' MWTP to avoid it provides a key metric for service design and pricing. We leverage an off-peak price discount implemented since 2016 and exploit exogenous variation in the crowdedness. The price discount, known as the Early-Bird Discount (EBD), gives a 30% discount to trips that originate from 16 stations on two suburban lines and start before 7 AM on workdays. A passenger chooses when to travel from the origin station (O) to the destination station (D), balancing three competing considerations. First, the passenger has an incentive to travel before the EBD cutoff to get a lower fare. Second, the passenger tries to avoid an overly crowded subway train. Third, the passenger has an optimal arrival time, and any deviation from that optimal time incurs a cost.

Instead of asking for passengers' subjective evaluations, we build a discrete choice model of travel demand that back out passengers' preferences for price, crowdedness, and deviation from the optimal schedule based on their observed choices. Crucially, the extent to which passengers are willing to pay a higher price for a less-crowded train indicates their WTP to avoid crowding in monetary terms. A key advantage of this setting is that crowding can be seen as a pure discomfort and is unlikely to be correlated with other important features that passengers may care about: Subway cars are otherwise identical, and we show that crowding does not lead to slower trains.

The main data used in the paper is the universe of trip-level records captured by smartcards in the Beijing Subway. The sample includes morning peak-hour trips around the cutoff that originate from the EBD stations as well as a set of comparison stations on two other suburban lines. The trip-level data only records the origin and destination stations and the corresponding time stamps when the passenger taps in and out of a subway station. To measure experienced crowding, we need to assign each passenger to a train at any given time. We split the observed travel time into in-train time as well as time spent in stations (e.g., walking in a station and waiting at the platform). We devise an iterative algorithm and jointly estimate the optimal route from O to D and the average access, transfer, and exit time at each station at any given time, with the in-train time imputed from the scheduled train timetable. This algorithm provides each passenger's precise location at any given time. Dividing the number of passengers in the same location by the available train floorspace at the time yields crowding measured by number of passengers per square meter. Subway trains during Beijing's morning rush hour are crowded, with an average measure of 3.31 passengers per square meter and the 90th percentile reaching 5 passengers per square meter. We assume passengers form correct crowding expectations based on the historical distribution of crowding in the same station pair at the same time of the day.

Despite the discomfort from crowding, passengers choose to travel in the same direction at the same time because they experience a common demand shock. For example, different employers may all adopt the same, conventional work hours, and commuters need to arrive at their workplaces on time. To break the endogeneity driven by correlated demand shocks, we construct an instrumental variable based on the concept of "accidental companion passengers" (ACPs). For a passenger i who is traveling from O to D at time t, an ACP is someone who travels from a different origin O' to a different destination D' at a different time t', but shares part of the trip with passenger i. We further restrict ACPs to those who take the subway infrequently and whose origins and destinations are sufficiently far from O and D, ensuring that the trips are responding to uncorrelated demand shocks. The instrumental variable for crowding is then constructed as the weighted sum of ACPs, where the weight is the overlapped travel time. The weighted ACPs is termed "accidental companion time" (ACT).

To incorporate the disutility of deviation from the optimal arrival time, we infer each passenger's unobservable optimal arrival time from her "coworkers," defined as passengers who used to arrive at the same destination around the same time, but live elsewhere and are unaffected by the EBD. Coworkers are identified using pre-EBD data from 2015, and the optimal arrival time for each passenger is then taken as the median arrival time of their coworkers.

We find that passengers have a substantial MWTP to avoid crowding. The average passenger is willing to pay about 6 cents RMB for each minute to reduce crowding by one passenger per square meter in the train. Consider the average trip in the sample that spends about 40 minutes in a train with average crowding of 3.31 persons per square meter, this amounts to 2.4 RMB, or about 46% of the fare, and the implied crowding externality created by the trip is about 10 RMB. The MWTP increases monotonically as crowding level rises, and high-income passengers are less sensitive to fare but derive larger disutility for crowding. The cost of rescheduling is also substantial and increases with income. A 15-minute deviation from the ideal travel time amounts to about two RMB.

We then investigate the welfare impacts of three alternative policies designed to reduce subway crowding: an optimal crowding tax, quantity control at optimal ridership, and a two-class (business vs standard) configuration. The optimal crowding tax is set at 5.5 RMB, which would more than double the baseline fare of 5.03 RMB. The higher tax-inclusive fare reduces total ridership by 28% and disproportionately affects low-income passengers. A quantity control policy slightly benefits low-income passengers but at a large cost to high-income users, resulting in lower overall system-wide welfare and larger externality on surface roads. In contrast, the two-class configuration allows high-income passengers to choose a pricier but less-crowded business class, while low-income passengers remain in standard class. This second-degree price discrimination delivers significant welfare gains for both high- and low-income groups and reduces external costs on surface road congestion.

Contributions to the literature: This paper makes several contributions to the literature. First, A large body of theoretical work suggests users' responses to prices, road congestion, as well as crowding within public system are important considerations in the design of transit provisions (e.g., Parry and Small, 2009; Coulombel and Monchambert, 2023). While there is a growing empirical literature on demand elasticity in public transit (e.g., Davis, 2021; Gu et al., 2023; Hahn et al., 2023) and road congestion, the study on within-system crowding is limited. Our paper is among the first to credibly estimate the disutility from crowding. We highlight that the subway setting in this paper presents important conceptual differences from the large literature on road congestion (e.g., Vickrey, 1969). While road congestion primarily generates externalities through lost time, crowding in public transit creates disutility in the form of discomfort.

Many of existing studies on crowding disutility rely on passenger surveys to elicit stated preferences (Douglas and Karpouzis, 2006; Lu et al., 2008; Li and Hensher, 2011; Wardman and Whelan, 2011; Haywood and Koning, 2015; Singh et al., 2023)., which may be subject to hypothetical bias and lack external validity. Some recent studies estimate passenger preferences based on observed choices using detailed trip-level records. For example, Hörcher et al. (2017) studies the Hong Kong Subway and Yap et al. (2020) uses a Dutch case study. Both studies model discrete route choices where some routes cost less time but are more crowded, while others are less crowded but take longer time to travel. Both studies estimate a "MWTP in time," which the authors later translate into monetary terms using rule-of-thumb levels of value of time. Instead, we build an economic model that incorporates several important dimensions of passengers' choices and provide the first revealed-preference estimates of passengers' willingness to pay to avoid crowding in monetary terms, leveraging the exogenous variation in price generated by the EBD. Moreover, we address the endogeneity of crowding by constructing a novel instrumental variable and estimate heterogeneity in the preference using a discrete choice model with random coefficients.

Second, the primary building block of a large class of structural travel models that features endogenous congestion is travelers choosing departure time. This class of models includes the bottleneck model (Vickrey, 1969; Arnott et al., 1993), the hydrodynamic traffic flow model (Lighthill and Whitham, 1955; Richards, 1956), the no-propagation model (Henderson, 1974; Chu, 1995), the bathtub model (Arnott, 2013), and the cell transmission model (Daganzo, 1994). The optimal arrival time is a crucial component of such models, but it is typically unobservable. We provide a novel way of measuring optimal arrival time by leveraging the network structure of the subway system and finding "coworkers" who are not affected by the EBD. We find that morning peak hour travelers have inflexible schedules, many are inframarginal, echoing some recent studies (e.g., Gu et al., 2023; Kreindler, 2024; Hall, 2024).

Third, we contribute to the study of optimal pricing under congestion externalities by being the first to estimate an optimal crowding tax. Our welfare analysis shows that, under heterogeneous disutility from crowding, second-degree price discrimination through a two-class (business vs. standard) configuration can illicit self-selection and result in a Pareto improvement. This adds to the long-standing discussion of price versus quantity controls (e.g., Weitzman, 2018; Li, 2018).

Lastly, although rich transit records are increasingly used in the literature (e.g., Gu et al., 2023), they usually contain only extry and exit time and stations, but cannot trace passengers. We develop a novel algorithm to impute real-time passenger locations, optimal routes, and in-train crowding. In addition, we construct "coworkers" to recover unobservable optimal arrival times. Together, these innovations allow us to capture both crowding and rescheduling disutility with greater precision.

The rest of the paper is organized as follows. Section 2 describes the background and the data. Section 3 introduces the demand model and describes the estimation and identification procedures. Section 4 describes how the main measures are constructed. Section 5 describes the sample and summary statistics. Section 6 presents the estimation results. Section 7 evaluates welfare impacts of alternative policy tools, and Section 8 concludes.

2 Background and Data

2.1 Background

Beijing is China's capital and the second-largest city. In the past two decades, it has built an extensive subway system. By 2019, the system consisted of 25 lines, over 400 stations, and an annual ridership of 3.8 billion. Subway has a particular advantage in long-distance commuting trips due to its fast speed and reliability. Excluding walking trips, the subway accounted for 15% of the commuting trips and about 40% of the passenger mileage in 2014 (Beijing Transport Institute, 2015). A typical subway trip is 15 km long and takes about 37 minutes, including accessing and waiting time.

The pricing of a subway ride is a step function of distance. A single-trip fare starts from 3 RMB for a ride under 6 km and increases by 1 RMB for additional 6 to 20 km, eventually reaching to 10 yuan for a ride between 92 and 112 km. Starting in December 2015, an early-bird discount (EBD) was applied in 16 stations on two suburban lines (Batong Line and Changping Line). A 30% discount is applied to the fare if a passenger enters one of those stations before 7 AM.

The 16 EBD stations are among the busiest during the morning rush hours. But they are not the only busy stations. The EBD was expanded to another eight stations on Line 6 by the end of 2016, which is outside of our sample's coverage. Other suburban lines that connect dense residential neighborhoods with the city center also have similar ridership patterns. We select 16 stations on Lines 6 and Line 5, which connects the northern suburbs to the city center. We show trips originating from the control stations experience similar crowding with those originating from the EBD stations.

2.2 Data

The Beijing Subway uses an electronic smartcard as the payment method. It is widely popular because it is easy to use and only card users qualify for discounts. By comparing the official statistics with our data, we estimate that between 90% and 95% of the subway trips were paid for by a smartcard around the time of the fare adjustment. The smartcard captures the entry and exit times and stations for each trip. Trips can be linked via an anonymized card number. The main data used in this paper is the universe of subway trips captured by smartcards and took place in about 20 weeks spanning between 2015 and 2016.

We geocode subway stations and collect the track distance and fare between each station pair from Beijing Subway's website. We collect train frequencies at each line and station. We also collect information on the types of passenger cars used in each line and the number of cars in each train, from which we obtain total floor space in a train. We do not separate seating areas from standing areas.

3 Model and Estimation

3.1 Demand Model

This section develops an empirical model of travel demand and crowding avoidance in the setting of urban rail transit. A potential passenger travels between a fixed origin-destination pair, captured by the entry and exit stations. The passenger chooses the departure time to maximize the expected utility, which depends on the fare, expected crowdedness during the trip, and the deviation from the optimal departure time.

In the demand model, a *market* is defined as an origin-destination (OD) station pair on a working day d. We focus on morning rush hours between 6:30 and 8:30 AM, around the EBD cutoff time at 7 AM. In each market, the passenger chooses a *product*, which is defined as the departure time t in 15-minute bins, measured by the tap-in time at the gate of the station of origin. Therefore, there are eight unique products in each market. The outside good in each market is trips to travel in the same station pair OD on day d during the morning rush hours but in other modes.

Passenger i's utility from riding the subway from station O to station D (the OD pair is henceforth denoted as j) on day d with departure time t is:

$$U_{ijdt} = \underbrace{\alpha_i P_{jt} + \beta_i \mathbb{E}[Crowd_{jdt}] + \rho_i |t + \mathbb{E}[TT_{jdt}] - t^{OA}_{jdt}| + \kappa_{jd} + \lambda_t + \xi_{jdt}}_{V_{ijdt}} + \epsilon_{ijdt}, \tag{1}$$

where P_{jt} is the charged fare. The fare varies by j as a step function of the track distance between O and D. It also varies by t within j in EBD stations, as passengers enjoy 30% off if they tap in the origin station before 7 AM. $\mathbb{E}[Crowd_{jdt}]$ is the expected crowding level in the subway for the trip. $\mathbb{E}[TT_{jdt}]$ is the expected travel time, t^{OA}_{jdt} is the optimal arrival time, and thus $|t + \mathbb{E}[TT_{jdt}] - t^{OA}_{jdt}|$ represents the rescheduling or deviation from the optimal arrival time. The rescheduling is measured in absolute values, which has two implicit assumptions. First, the penalty from being late or early than the optimal arrival time is assumed to be symmetric; second, the penalty is assumed to be linear in the time deviation. Those assumptions are simplifying as the cost of resecheduling is not the focus of our estimation, and they can be easily relaxed.

κ_{jd} is the common utility of completing trip j on day d, which is the common value from participating in the market and can be captured by a set of OD-pair-by-date fixed effects. λ_t is the common value of departing at time t across all markets and products, which can be captured by a set of time bin fixed effects. ξ_{jdt} is the product-specific demand shock that is unobserved by the econometrician.

The marginal utility from the fare, expected crowdedness, and the deviation from the ideal arrival time is denoted by α_i, β_i, and ρ_i, respectively. The preferences are heterogeneous and differ by income. The construction of income distribution of subway riders in each OD pair is described in Section 4.1. V_{ijdt} captures the deterministic part of utility individual i derives from choosing the product jdt. Each passenger also draws a mean-zero stochastic utility for choosing the product, ϵ_{ijdt}, which is drawn $i.i.d$ from a type-I extreme value distribution.

3.1.1 Standard Logit Model

When there is no heterogeneity in preference, the type-I extreme value distribution assumption translates the problem into a simple multi-nominal logit. Assuming the utility of choosing the outside good to be zero, the probability of choosing departure time t in market jd can be written as:

$$S_{jdt} = \frac{exp(V_{ijdt})}{\sum_{t' \in \{T,0\}} exp(V_{ijdt'})}, \tag{2}$$

where $\{T, 0\}$ is the choice set with T indicating the set of 15-minute bins and 0 the outside good. S_{jdt} is the market share of product t. The market share can be inverted as (Berry, 1994):

$$log(S_{jdt}) - log(S_{jd0}) = \alpha P_{jt} + \beta \mathbb{E}[Crowd_{jdt}] + \rho|t + \mathbb{E}[TT_{jdt}] - t^{OA}_{jdt}| + \kappa_{jd} + \lambda_t + \xi_{jdt}, \tag{3}$$

where S_{jd0} is the market share of the outside good, which will be absorbed by the market fixed effects κ_{jd}.

The expectation of crowding level and travel time can be approximated based on historical records (Cook and Li, 2025). The optimal arrival time t^{OA}_{jdt} is inferred by identifying coworkers – passengers who arrive at the same destination at the same time in a pre-period but live in different parts of the city. Using d_{-1} to denote days prior to day d, equation 3 is transformed to:

$$log(S_{jdt}) = \alpha P_{jt} + \beta Crowd_{jt,d_{-1}} + \rho|t + TT_{jt,d_{-1}} - t^{OA}_{jdt}| + \kappa_{jd} + \lambda_t + \xi_{jdt}. \tag{4}$$

How $Crowd_{jt,d-1}$, $TT_{jt,d-1}$ and t^{OA}_{jdt} are measured is described in Section 4.1. α, β, and ρ are the marginal utility of price, crowding, and departure from the optimal schedule, respectively. We expect them to all have negative values. Having variation in prices and crowding allows for the estimation of the willingness to pay (WTP) to avoid crowding in dollar terms. The WTP for one unit reduction in crowding is β/α; the WTP for one minute reduction in the deviation from the optimal arrival time is ρ/α. The identification challenge is that ξ_{jdt} is potentially correlated with P_{jt} and $Crowd_{jt,d_{-1}}$. The identification strategy involves using a control set of stations and an instrumental variable for crowding, which we introduce in Section 3.2.

3.1.2 Random Coefficient Model

Passengers with different characteristics may have heterogeneous preferences. They may have different price elasticity and derive different levels of disutility from crowding and rescheduling. We are particularly interested in preference heterogeneity by income. We allow α, β and ρ to differ by individual and specify $\alpha_i = \alpha_0 + \alpha_1 I_i$, $\beta_i = \beta_0 + \beta_1 I_i$, and $\rho_i = \rho_0 + \rho_1 I_i$, where I_i is the simulated passenger income described in Section 4.1.[1] I_i is recentered at the sample average and thus has a mean of zero.

The utility function becomes:

$$\begin{aligned} U_{ijdt} = &\underbrace{\alpha_0 P_{jt} + \beta_0 Crowd_{jt,d_{-1}} + \rho_0 |t + TT_{jt,d_{-1}} - t^{OA}_{jdt}| + \kappa_{jd} + \lambda_t + \xi_{jdt}}_{\delta_{jdt}} \\ &+ \underbrace{\alpha_1 I_i P_{jt} + \beta_1 I_i Crowd_{jt,d_{-1}} + \rho_1 I_i |t + TT_{jt,d_{-1}} - t^{OA}_{jdt}|}_{\mu_{ijdt}} + \epsilon_{ijdt}. \end{aligned} \tag{5}$$

We use δ_{jdt} to denote the product-level utility that is common to all passengers and μ_{ijdt} to denote the passenger utility due to income levels.

The random coefficient model does not have a simple closed-form expression for market share. The predicted market share of tap-in time t in market jd can be approximated using Monte Carlo integration (Berry et al., 1995).

$$s_{jdt} \approx \frac{1}{n_{jd}} \sum_{i=1}^{n_{jd}} s_{ijdt} = \frac{1}{n_{jd}} \sum_{i=1}^{n_{jd}} \frac{\exp(\delta_{jdt} + \mu_{ijdt})}{\sum_{t' \in \{T,0\}} \exp(\delta_{jdt'} + \mu_{ijdt'})}, \tag{6}$$

where n_{jd} is the number of passengers in market jd and s_{ijdt} is the probability of passenger i choosing tap-in time t in market jd.

Denote the market share observed from the data as S_{jdt}, we can calculate δ_{jdt} numerically by the fixed-point iterations:

$$\delta^{h+1}_{jdt} = \delta^{h}_{jdt} + log(S_{jdt}) - log(s_{jdt}) \quad \text{for } h = 0, \dots, H.$$

The unobservable demand shock can be written as

$$\xi_{jdt} = \delta_{jdt} - (\alpha_0 P_{jt} + \beta_0 Crowd_{jt,d_{-1}} + \rho_0 |t + TT_{jt,d_{-1}} - t^{OA}_{jdt}| + \kappa_{jd} + \lambda_t) \equiv \omega_{jdt}.$$

Denote the vector of the parameters as θ and a set of instrumental variables as Z_{jdt}. The GMM estimate is:

$$\hat{\theta} = \arg\min \omega_{jdt}(\theta)'(Z_{jdt})\Phi^{-1}(Z'_{jdt})\omega_{jdt}(\theta). \tag{7}$$

where Φ^{-1} is the optimal weight matrix for the GMM estimation.

[1] Allowing for random coefficient requires additional instrumental variables, and many potentially weak instruments lead to biased estimates. We adopt a parsimonious model to focus on the key components of the model. Note that preferences are not allowed to vary along unobserved characteristics.

3.2 Identification and Estimation

3.2.1 Challenges to Identification

Challenging the consistent estimation of preference parameters comes from the variation in price generated by the EBD and the potential endogeneity between the expected crowdedness and unobserved product characteristics ξ_{jdt}. Below, we discuss the challenges for each case.

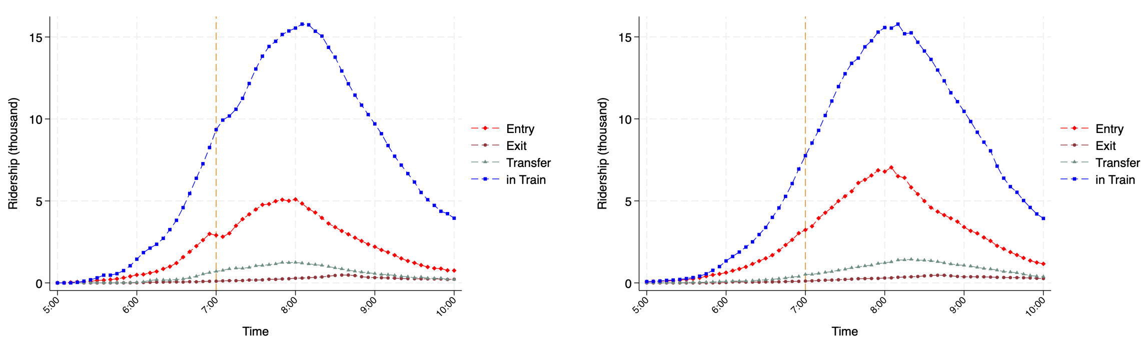
The listing fare for a subway ride between the station pair *OD* is determined by the track distance between the two stations. The EBD creates time-variation in price around the time cutoff. The choice of the time cutoff (7 AM) is arbitrary during a set of possible times during the morning peak hours. The morning system load does not peak until around 8:15 AM. The discount rate, at 30%, can also be thought of as arbitrary. In fact, the discount rate was raised to 50% by the end of 2016, which is after the end of our sample period. While the specific parameters of the EBD are unlikely to be correlated with the product value due to unobserved characteristics, two concerns arise when using the price variation generated by the EBD. First, the choice of stations to implement the EBD may not be arbitrary. The 16 EBD stations are on two suburban lines, and are among the busiest during morning rush hours. Note that we control for market fixed effects (κ_{jd}) and the preference for price is identified using only time variation created by the EBD. Second, the discount is a fixed percent of the listing fare. While there is variation in price across different *OD* pairs after controlling for departure time fixed effect, λ_t, the variation is driven by differences in price *levels*. If, for example, log price is included instead of price level, α is not identified with κ_{jd} and λ_t controlled for. In other words, with only EBD stations in the sample, the model is identified via a specific functional form. To address this problem, we introduce a set of control stations that have similar ridership patterns but were not selected for the EBD during the sample period. We introduce those comparison stations in Section 3.2.2.

Expected crowdedness during the trip, proxied by historical crowding levels in the same *OD* pair j with the same departure time t on previous days, $Crowd_{jt,d_{-1}}$, can be correlated with the unobserved value of traveling in the same *OD* pair with the same departure time t on day d. While expected crowding based on historical patterns is unrelated to the unobserved demand shocks in the specific day, crowding patterns are expected to be highly similar in the same time on different days. A subway ride in the certain *OD* pair at a certain time t is crowded not because passengers enjoy crowded subway cars, but because taking the specific trip at the specific time has particularly high value *despite* high crowding levels. Those product (jdt)-specific utility components include the joint spatial distribution of home and work locations, as well as the clustering of optimal arrival times. For example, if a large share of residents living near the station O work near station D, the commute (combining the subway trip and the access time to the workplace) takes about 30 minutes, and workers need to arrive at work by 9 AM while there is little return to arrive early, we would expect to see many tapping in at station O at around 8:30 AM despite the expected high crowdedness of the subway. This generates a positive correlation between ξ_{jdt} and $Crowd_{jt,d_{-1}}$. Ignoring this correlation would bias β towards zero or even positive. We introduce an instrumental variable for crowdedness, building on the concept of *accidental companion passengers*, whose trips start from a different origin, end in a different destination and are driven by unobserved demand shocks that are arguably uncorrelated with the demand for product jdt under consideration, but nevertheless share the same subway car for a subset of the trip jdt and contribute to $Crowd_{jdt}$. The construction of the instrumental variable and discussions for its relevance and validity are described in Section 3.2.3.

3.2.2 Control Stations

The EBD was implemented at 16 stations on two busy suburban lines at the end of 2015. 11 of those stations are on the Batong Line, which connects the populous eastern suburbs with the central business district; the other five are on the Changping Line, which connects the dense northern suburbs to other loop lines and lines that connect to the city center. Those EBD stations are at the farthest ends of the two subway lines. They are among the busiest stations during weekday morning peak hours. Figure 1 shows that on average, more than 5,000 passengers tap in at one of the 16 stations between 8 and 8:05 AM. The vast majority of those passengers are downtown-bound. Assuming the frequency of the trains during morning peak hours is every three minutes, there will be about 180 passengers waiting on the platform when each train pulls into a station (5000/16/5*3). Those trips are typically long, as suggested by the very low numbers of exits. Therefore, trains quickly get crowded. At around 8:15 AM, about 15,000 passengers who entered one of the EBD stations are still somewhere in the system.

Figure 1: Ridership by Time, EBD and Control Stations



(a) EBD Covered Stations (b) Control Stations

Notes: The graphs show the number of entries, exits, transfers, and in-train passengers by 5-minute intervals during morning hours in EBD stations (Panel a) and control stations (Panel b).

To qualify for a good control group, candidate stations need to exhibit similar patterns in ridership over time. Ideally, they should have similar characteristics with the EBD stations along unobserved dimensions as well (e.g., the composition of unobserved passenger characteristics). We select 16 stations on the far ends of two other suburban lines, Line 5 and Line 6, as control stations.[2] As shown below, these control stations exhibit comparable patterns in entries, exits, transfers, and the number of over time. In addition, the control stations are likely to be comparable along dimensions that are not directly measurable. Line 6 is parallel to the Batong Line, and the eight stations in the control group implemented EBD at the end of 2016. Line 5 cuts through the city center from north to south. Its northern branch is parallel to the Changping Line and passes through dense residential neighborhoods. Though not covered by the EBD till this day, the eight stations we choose as the control group are in constant debate to be included in the EBD.

Figure 1 panel (b) reports measures of ridership at the 16 control stations in five-minute bins. Similar to the EBD stations, entries and in-train passengers at control stations peak at around 8:15 AM. Trips from control stations are also long, and the number of exits remains low during the sample time window. Though not necessary for identification, these measures are also comparable in levels. The correlation coefficients between EBD and control stations at the same five-minute bin are high for the number of entries, exits, transfers, and the number of passengers in trains. Notably, while the numbers of entries have a visually salient kink around the EBD cutoff at 7 AM — some trips originally scheduled for some time after 7 AM are moved to some time before to qualify for a discounted fare — there is no such kink for control stations.

3.2.3 Instrumental Variable for Crowding

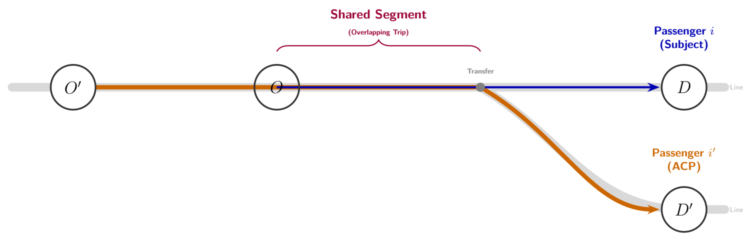
Crowding in the subway may be endogenous. It is reasonable to think the expected crowding, $Crowd_{jt,d_{-1}}$, which is based on historical levels of crowdedness in trips in the same OD pair leaving at the same time t, is positively correlated with the unobserved demand shock, ξ_{jdt}. Crowdedness for trips in the same jt are highly correlated across different days. The subway is crowded precisely because many other passengers decide to take the same subway trip at the same time, driven by common unobserved demand shocks, ξ_{jdt}. Ignoring such positive shocks will bias the estimated preference for crowdedness upward.

We propose a novel IV for crowding based on the concept of *accidental companion passengers* (ACP). The ACPs are passengers who tap in at a different origin station O', tap out at a different destination station D', but nevertheless share the same train with a passenger in trip $OD-t$ for a subset or entire of his trip. Figure 2 illustrates the idea behind the instrument. Consider a passenger i who travels from O to D at tap-in time t. His experience of crowdedness in the subway car is partly due to passenger i', who starts at the upstream station O' at t', boards the same train, and travels to

[2]The estimation model controls for market (OD pair-by-date, or jd) fixed effects, so the ridership in the control stations does not need to be comparable in *levels*. However, the temporal patterns need to be comparable as the model explores variation in crowdedness over time for the same station pair-day. Alternatively, we could construct "synthetic" control stations based on statistical patterns.

D'. i and i' share the same train for a segment of the trip. ACPs may also board at a station on a line other than the one station O is located, or exit at a station on a different line from station D (as is passenger i').

Figure 2: Accidental Companion Passengers



ACPs in a trip $O'D'$ (j') on day d at time t' is affected by the unobserved demand shock $\xi_{j'dt'}$. Therefore, the validity of the IV hinges on the assumption that $cov(\xi_{j'dt'}, \xi_{jdt}) = 0$. While this assumption is not directly testable, we improve the credibility of the IV by choosing ACPs whose demand for travel is unlikely to be correlated with ξ_{jdt}. Several additional refinements are imposed on ACPs. First, we only include O's and D's that are sufficiently far from O and D, respectively. In the baseline, we include ACPs to a destination station D' that is at least eight kilometers away from D. We exclude ACPs from an origin station O' that is downstream from O, as they may be directly affected by the choice of jdt trips. We test the robustness of the result when we vary the size of the hole between O and O', and between D and D'. Second, travels in jdt may be ACPs for other trips. To break down this reverse causality, we construct a refined measure of ACPs that are based on infrequent subway riders, defined as those who take less than four trips per week. The intuition behind this refined ACP measure is that frequent (such as peak-hour commuters with a regular schedule) and infrequent users (such as retirees and leisure travelers) respond to demand shocks drawn from uncorrelated distributions. In our sample, infrequent users account for over half of the passengers but less than 10% of the total trips. Using ACPs in $O'D'$ pairs that are far from OD and using infrequent riders who account for a small share of the overall ridership potentially weakens the power of the instrument. But instrument relevance is testable as the power of the first stage, which we report in the next section.

We further incorporate the duration of overlap (in travel time) between the ACP and the trip in question, and define the *accidental companion time* (ACT) as the weighted sum of ACPs, where the weight is given by the overlapped travel time. We employ ACT as our primary instrumental variable for crowding in the analysis. Because expected crowding is constructed based on historical levels, we measure ACPs and ACTs using the corresponding days in the past.

We need at least four additional instruments to estimate the random coefficient model in order to estimate the random coefficients of the preference for price, crowdedness, rescheduling and the intercept. The first is the interaction between an indicator for the EBD stations and an indicator for the tap-in time before 7 AM (the EBD cutoff). This variable, with control stations that have similar ridership patterns but are not covered by the EBD, uses exogenous variation introduced by the EBD, which shifts passengers' optimal departure times and generates exogenous variation in crowding. The other three additional instruments are interactions between the EBD station indicator and linear, quadratic, and cubic time trend. The validity argument is the same, and they are correlated with crowding through serial correlation.

4 Measurement

This section describes the sample and key measures such as crowdedness, distribution of optimal arrival time, and the income distribution among passengers in each OD pair.

4.1 Construction of Variables

4.1.1 Optimal Route

To measure crowdedness in the subway, we need to assign each passenger to a train. There are two key challenges to this task. First, we only observe passengers' tapping-in and tapping-out stations and times, not which trains they board. We also do not observe trains' locations, or the real-time crowdedness on platforms or inside trains.[3] Second, with more than 20 lines and 400 stations, the Beijing Subway in 2016 was a complex network, and there is often multiple possible routes between a given OD station pair. We need to determine which route each passenger takes.

We assume that passengers take the route with the shortest travel time, and identify this optimal route using observed data. Specifically, we define a route $\mathcal{P}_{jdt}$ as an ordered sequence of stations and their associated times:

$$\mathcal{P}_{jdt} = \{S^O(t^O), S_1(t^1), ..., S_n^R(t^n), ..., S^D(t^D)\},$$

where S^O and S^D denote the entry and exit stations, respectively, and S^R denotes a transfer station, if applicable. We let $\mathcal{R}_{\mathcal{P}\,jdt}$ represent the set of transfer stations along route $\mathcal{P}_{jdt}$.

The total travel time $TT^{\mathcal{P}}_{jdt}$ for route $\mathcal{P}_{jdt}$ consists of four components:

$$TT^{\mathcal{P}}_{jdt} = ST^{\mathcal{P}}_{jdt} + AT^{O}_{t^O} + \sum_{R\in\mathcal{R}_{\mathcal{P}\,jdt}} RT^{R}_{t^R} + ET^{D}_{t^D}$$

where $ST^{\mathcal{P}}_{jdt}$ is the in-train time, defined as the sum of train travel times between adjacent stations;

[3]Most trains in the Beijing Subway are equipped with real-time location devices, which are used to display precise estimated arrival times at stations. Furthermore, platforms and passenger cars have cameras, which can potentially be used to derive measures of crowdedness. However, we do not have those data.

$AT^O_{t^O}$ is the access time, defined as the time between tapping in and boarding the train; and $ET^D_{t^D}$ is the exit time, defined as the time between debarking the train and tapping out. Some routes require transferring, and we use $RT^R_{t^R}$ to denote the transfer time at transfer station R when arriving at the station at time t^R.

For each route, we observe the tapping-in time t^O and tapping-out time t^D. We calculate the in-train time $ST^{\mathcal{P}}_{jdt}$ based on the arrival times of the first and last scheduled trains for each line at each station published by the Beijing Subway, assuming trains always operate on time. Since access, exit and transfer times are unobserved, we begin by assigning initial guess values to them. These initial values allow us to back out arrival times at transfer stations t^R and calculate the total travel time $TT^{\mathcal{P}}_{jdt}$. Using the calculated total travel time, we apply Dijkstra's algorithm to identify the optimal route $\mathcal{P}^O_{jdt}$, with the total travel time of the optimal route denoted as OT_{jdt}.

To get better estimates of the access, transfer, and exit times, we estimate the following equation:

$$\hat{TT}_{jdt^O} = \hat{ST}_{jdt^O} + AT^O_{t^O} + \sum_{R\in\mathcal{R}} RT^R_{t^R} + ET^D_{t^D} + \varepsilon_{jdt^O} \tag{8}$$

where variables with a $\hat{hat}$ are observed in the data. Specifically, $\hat{TT}_{jdt^O}$ is the *median* total travel time in actually observed trips from the smartcard data between OD pair (j) with tapping-in time t^O, and $\hat{ST}_{jdt^O}$ is the in-train time in the optimal route imputed from the scheduled train timetable. We estimate $AT^O_{t^O}$, $ET^D_{t^D}$, $RT^R_{t^R}$, and the associated arrival time t^R using the Ordinary Least Squares (OLS), pooling across the entire subway network. Time is discretized into 5-minute bins. There is enough degree of freedom to flexibly identify the access, transfer, and exit times specific for each station and time bin. The estimation is weighted by the number of trips in the (jdt^O) bundle.

Table 1: In-Station Time Estimation

	Estimated In-Station Time (min)				
	Mean	SD	P10	P90	N
Entry	3.50	1.84	1.36	5.34	33,436,600
Exit	2.00	0.90	0.89	3.15	33,436,600
Transfer_pool	4.43	1.42	2.61	6.23	34,782,743
Transfer_1st	4.44	1.44	2.57	6.19	22,872,454
Transfer_2nd	4.51	1.56	2.40	6.30	9,282,034
Transfer_3rd	4.10	2.02	1.60	6.89	2,252,395
Transfer_4th	4.27	3.22	2.01	7.55	337,176
Transfer_5th	2.49	4.30	-5.46	7.65	36,284
Transfer_6th	3.05	0.90	2.88	2.88	2,400

Notes: This table reports the estimated in-station time using Equation 8. In-station time is separated by entry time, exit time (with the mean imposed at 2 minutes), and transfer time.

We recalculate the optimal route choice and in-train time for jdt^O trips using the estimated $\tilde{AT}^O_{t^O}$,

$\tilde{ET}^{D}_{t^D}$, and $\tilde{RT}^{R}_{t^R}$. Note that we assume the operating speed of the train does not vary by t, but the optimal route may vary over the course of the day due to time-varying crowdedness in the stations, so the in-train time cost, $\hat{ST}_{jdt^O}$, comes with a t^O subscript. We iterate this process until the estimates of $\tilde{AT}^{O}_{t^O}$, $\tilde{ET}^{D}_{t^D}$, and $\tilde{RT}^{R}_{t^R}$, and $\hat{ST}_{jdt^O}$ (corresponding to the optimal route choice) converge. In practice, the optimal route for a given OD pair rarely varies by the time of the day t^O, so we assign the most common optimal route over the course of the day to each OD pair.

Given a trip's optimal route, Equation 8 faces a colinearity problem, and all time segments cannot be estimated jointly. To see that, notice that $\hat{TT}_{jdt^O}$ is observed from the data, in-train time $\hat{ST}_{jdt^O}$ is determined by the train schedule, with the error ε_{jdt^O} being mean-zero, in-station time costs, $AT^{O}_{t^O}$, $RT^{R}_{t^R}$, and $ET^{D}_{t^D}$, add up to a constant equal to the mean total travel time minus in-train time. To break the colinearity, additional restrictions need to be added to in-station time costs. We restrict the exit time to have a mean of two minutes.

Using Equation 8, we estimate the entry and exit time for each station, and the transfer time for each transfer station. In the baseline estimation, in-train times are the same within for the same OD and in-station times are the same for the same station-purpose combination. In other words, there is no t subscript in Equation 8.

Table 1 summarizes the estimated time segments by category. Entry time, from tapping into the origin station to boarding the train, takes an average of 3.5 minutes. Estimated entry time varies across stations, probably due to different station configurations and train frequency, and has a standard deviation of 1.84 minutes. Exiting takes an average of two minutes, which is pre-imposed. With an average of 4.43 minutes, transfer takes the longest time among all three in-station segments. Some transfer stations cost more than 6 minutes on average. Finally, the last column indicates the number of effective observations used to estimate each type of in-station time cost. Because all trips have an entry station and an exit station, the number of observations used to estimate those two types equals the total number of trips, or some 33.4 million. The average trip a little more than one transfer, so transfer time is estimated out of 34.8 million trips. Given the estimates in Table 1, we set the shares of access, exit, and transfer times in the total in-station time to be 5:3:8.

The algorithm to determine the optimal route imposes several implicit assumptions. First, the speed of the train is assumed to be the same at any t. In particular, we do not allow the speed to slow down as the train gets crowded. During the study period, the majority of the Beijing subway system was brand new. It was well equipped with the most advanced control system. While the passenger cars could get crowded, it was reported to be running precisely on time. Stopping time at stations is computer controlled. Most stations have double-layered doors to prevent accidents. Unlike some older systems in western countries, delay is rare in the Beijing Subway. We present in Section 5.2 that the relationship between crowding and time cost is negligible. Second, while the algorithm is based on the median passenger, we assume all users take the same route. In particular, we assume passengers do not take an alternative route as the shortest-time route gets crowded. Section 5.2

provides strong evidence that taking a detour to avoid crowding is rare.

Lastly, not all passengers traveling from O to D with tapping-in time t^O arrive at the same time. There is a distribution of total travel time along the optimal route. However, as we show in Section 5.2, the distribution is tight. We assume that trains are always on time and that the in-train time for each $OD - t$ is identical, and we attribute the observed distribution in travel time to idiosyncratic factors and assign them to in-station time costs. For example, some passengers walk faster than others; some passengers are lucky to catch a departing train while others may be less so; passengers may delay their trip by visiting a convenience store or vending machine inside the fare gates. While assuming all passengers take the same optimal route, we treat the sources of time cost variation as agnostic and proportionately scale up or down each segment of the median passenger's time cost to fit the observed time cost for each passenger in trip (jdt^O). For example, suppose the imputed time cost of trip (jdt^O) is 40 minutes, with 10 minutes access time, 5 minutes transfer time, 20 minutes in-train time, and 5 minutes exit time. Suppose we observe an actual trip in (jdt^O) that costs 44 minutes (10% more than the imputed time), we then assign 12 minutes for access, 6 minutes for transfer, 6 minutes for exit, while the in-train time remains at 20 minutes. With the optimal route imputed for all trips, a trip r can be summarized as a (OD, t^O, t^D) bundle. That is, a subway ride can be summarized as the origin and destination stations as well as tap-in and tap-out times).

4.1.2 Train Assignment and Crowding

We do not directly observe trains. To allocate a passenger to a train at any given time, we first define a "segment", s, as the train tracks between an adjacent station pair and the associated scheduled travel time. Using the optimal route and access, in-train, transfer, and exit times calculated in the previous subsection, we can determine the exact location of trip $r = (OD, t^O, t^D)$ at any time $t \in [t^O, t^D]$. Summing over real-time locations of all passengers for any 5-minute time bin τ yields the number of passengers in segment s at τ, which we denote as N^s_τ.[4]

We focus particularly on in-train crowding.[5] Let A be the floor area of a passenger car and M be the number of cars in a train[6], the in-train crowding experienced by passenger i in time bin τ,

[4]If multiple segments fall within a single 5-minute bin, we calculate the ridership of each segment weighted by its proportion of segment-specific travel time within that bin. If a segment spans multiple 5-minute bins, we partition it into sub-segments, each assigned to the corresponding consecutive 5-minute bin.

[5]Passengers may also derive negative utility from crowded subway stations. The optimal route and train assignment algorithms also allow us to calculate the number of passengers in the origin, destination, and transfer stations at any given time. Floor (or platform) areas of each station are available from Beijing Subway's website. We focus on in-train crowdedness because it is arguably the most important dimension. Including crowdedness in other trip segments is possible, but is likely to bring additional challenges to identification because (1) additional instrumental variables are needed to identify those crowdedness and their random coefficients and (2) in-train crowdedness and in-station crowdedness are for a given trip likely to be highly correlated, making separate identification challenging. Leaving in-station crowding out of the analysis does not bias our estimates as long as the instrumental variable is uncorrelated with in-train crowding.

[6]Each train in our sample consists of six passenger cars that each measures at 19 meters long and 2.8 meters wide, providing a total passenger floorspace of about 320 square meters. We do not consider floorspace taken up by seats and

located in segment s, is given as:

$$k_{ijd\tau} = \frac{N_\tau^s}{A \cdot M} \tag{9}$$

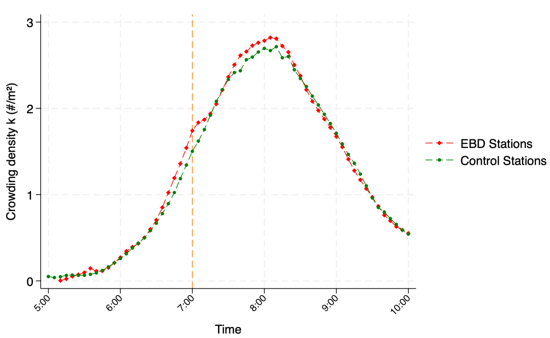
$k_{ijd\tau}$ has the unit of number of persons per m^2. The *total crowdedness* (K_r) and *average crowdedness* (k_r) experienced by the passenger during the entire trip $r = (OD, t^O, t^D)$ are calculated using the following two equations, respectively:

$$K_r = \sum_{\tau \in [t^O, t^D]} k_{ijd\tau} \tag{10}$$

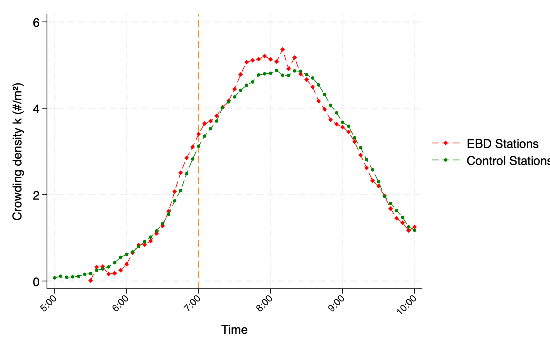
$$k_r = \frac{K_r}{t^D - t^O} \tag{11}$$

The unit for the total crowdedness is the number of passengers per square meter times minute (mins · persons/sqm). The unit for the average crowdedness is the number of passengers per square meter (persons/sqm). K_r is our main measure of crowdedness.

Figure 3: Crowding in Downtown Bound Trains



(a) Average across Stations



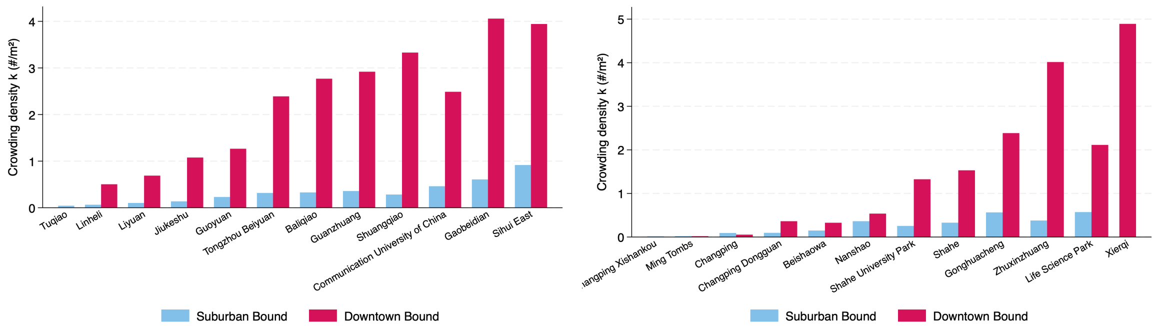
(b) The Most Downstream Stations

Notes: This figure depicts downtown bound "in-train" crowding (person/m^2) at each station over time. Panel a presents the average crowding across stations, while Panel b narrows down to the crowding of the most downstream stations from each line.

The left panel of Figure 3 shows the average in-train crowdedness during morning rush hours in downtown-bound trains passing through EBD and control stations. The treated and control stations have similar levels of crowdedness and exhibit similar patterns over the course of the morning rush hours. The average crowdedness is about 1.5 persons per square meter at around 7 AM, peaks at around 2.8 persons per square meter at 8:15 AM, and declines to 0.5 persons per square meter at 10 AM. The line representing EBD stations almost always overlays with that representing control stations except for the window around the EBD cutoff time of 7 AM, where the bunching behavior in EBD stations creates a hump in crowdedness. The graph on the right shows the crowdedness in

other on-train equipment. We also do not model the probability of finding a seat or potentially different utilities from sitting and standing. Using smartcard data from the Hong Kong subway similar to ours, Hörcher et al. (2017) presents an algorithm to determine the probability of finding a seat.

Figure 4: In-Train Crowdedness in EBD Lines by Direction



(a) Batong Line (b) Changping Line

Notes: The graphs show the crowding density k of subway trains that pass each EBD station on the two EBD lines: Batong Line in panel (a) and Changping Line in panel (b). In each panel, EBD stations are listed from the most upstream on the left to the most downstream on the right.

trains passing the most downstream stations in our sample. Data are from only four stations for this graph (one in each line, with two EBD lines and two control lines), so the lines are less smooth. Nevertheless, the crowdedness is again similar between the two EBD stations and the two control stations. With passengers boarded from upstream stations, trains passing the most downstream stations are significantly more crowded. At around 8:20 AM, the in-train crowdedness peaks at around five persons per square meter in the two EBD stations.

The stations in the sample are at the far ends of suburban lines, most passengers travel in the downtown during morning rush hours. Figure 4 shows that downtown-bound trains are far more crowded than suburban-bound trains in the two EBD lines. Trips in both directions qualify for the EBD and are in our sample. The substantial difference in crowdedness between trips in different directions allows us to identify the potentially varying marginal utility at different crowding levels.

Finally, we convert crowding measures to be used in the choice model. In the main specification in Equation 4, passengers traveling between OD (j) choose tap-in time t in 15-minute time bins based on historical crowding levels, $Crowd_{jt,d_{-1}}$. We construct $Crowd_{jt,d_{-1}}$ by taking average over the k_r of trips in the same OD pair and departure time bin t observed in the previous four weekdays.[7] Table 2 reports regression results predicting crowding with varying sets of fixed effects. The high explanation power, with R^2 values ranging from 0.65 to 0.99, indicates that passengers' crowding experiences can be reliably approximated using historical averages for the same OD pair and departure time bin.

4.1.3 Accidental Companions

We use accidental companion time (ACT) to instrument for crowding experienced by passengers in (OD, t) trips. ACT is defined as the weighted sum of accidental companion passengers (ACPs),

[7]Recall that our data consist of non-continuous weeks in 2015 and 2016. We exclude dates for which there are less than four weekdays within 60 days prior.

where the weight is the shared travel time.

Specifically, for passenger i's trip (OD, t), we define an ACP as passenger i' who travels between $O'D'$, where $O \neq O'$ and $D \neq D'$, but whose trip overlaps with passenger i for $m^{i,i'}$ minutes. We require that O's are in the same line as O but in the upper stream, and D's are at least 8 kilometers away from D. We are more restrictive on D's being sufficiently far away from D for two reasons. First and most importantly, the potential endogeneity concern is correlated demand shocks. They are more likely to be driven by the common destination (e.g., employers near the exit stations have similar deadlines for check-ins) rather than by the common origin. Second, trips in the sample are selected by origin stations (totaling 16 of them), and they are at the edges of suburban lines. Requiring O's to be far from O will yield a small number of ACPs. To further mitigate concerns about correlated demand shocks, we restrict ACPs to infrequent users, defined as those who take less than four trips during the week. Section 6 presents robustness results under ACPs defined with varying distances between O and O', D and D', and type of passengers.

Let I' be the set of infrequent ACPs, the ACT for passenger i is then given by:

$$ACT_{ijdt} = \sum_{i' \in I'} m^{i,i'} \tag{12}$$

Because expected crowding is based on historical crowding levels, the number of ACPs and ACT are constructed using trips in the same historical dates.

Table 2: Predictability of Crowding with Varying Fixed Effects

Panel A	Crowding K ($\# \cdot min/m^2$)			
	(1)	(2)	(3)	(4)
Constant	82.8767	82.8401	83.1659	84.2793
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
R^2	0.8301	0.9876	0.9916	0.9946
Panel B	**Crowding per minute k ($\#/m^2$)**			
	(1)	(2)	(3)	(4)
Constant	2.2845	2.2908	2.3477	2.4314
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
R^2	0.6511	0.9812	0.9870	0.9916
Fixed Effects				
OD	X			
OD-time		X		
OD-day of week-time			X	
OD–quarter-day of week-time				X

Notes: This table shows the predictability of crowding K (Panel A) and crowding density k (Panel B). The constant term indicates the average value of the dependent variable in the omitted group.

Table 3: Predictability of Ridership with Varying Fixed Effects

Panel A	# of Trips from Frequent Riders			
	(1)	(2)	(3)	(4)
Constant	1.8467	1.8467	1.8521	1.8628
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
R^2	0.6804	0.9342	0.9449	0.9673
Panel B	**# of Trips from Infrequent Riders**			
	(1)	(2)	(3)	(4)
Constant	0.2424	0.2424	0.2402	0.2344
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
R^2	0.4317	0.5226	0.6281	0.7502
Fixed Effects				
OD	X			
OD-time		X		
OD-day of week-time			X	
OD−quarter-day of week-time				X

Notes: This table shows how variation in ridership can be predicted by a varying set of fixed effects. Each observation is an OD-date-time bin. The depedent variable in Panel A is the number of trips from frequent riders, defined as those who have at least four trips in the week. The depedent variable in Panel B is the number of trips from infrequent riders. The constant term indicates the average value of the dependent variable in the omitted group.

Crowding and trip volumes and largely predictable. Table 2 shows that OD and time bin alone explain the variation in experienced crowding almost perfectly (with an $R-$squared over 0.98). In terms of ridership composition in a given OD pair at time bin t on date d, Table 3 shows that ridership from frequent riders is highly predictable: OD-by-time bin fixed effects explain more than 93% of the variation. On the other hand, trips from infrequent riders are somewhat less predictable, although they account for a much smaller share of trips. However, even for infrequent riders, accounting for OD pair and time bin has an $R-$squared over 0.5, and further accouting for the season and day of week raises the $R-$squared to be over 0.75.

While the instrumental variable is based on expected crowding created by infrequent riders from upstream stations, it does not require downstream passengers to rationally predict the number of ACPs. While the predictability of infrequent riders is not particularly low, the predictability and magnitude of ACPs due to infrequent users feeds into the strength of the first stage, which is empirically testable.

4.1.4 Optimal Arrival Time

To estimate the disutility from rescheduling, we need to measure the optimal arrival time (t^{OA}_{jdt}) in Equation (1) for the observed (jdt) trips. This variable, however, is fundamentally unobservable. Observed arrival times may not be ex ante optimal as passengers weigh trade-offs among rescheduling costs, temporal fare differences due to the EBD, and expected crowdedness at different departure times. Historical arrival times of the same passenger are also not good proxies, since our data do not allow us to link many individual passengers historically and a passenger's optimal arrival time may vary.

We therefore infer the optimal arrival time of passengers in (jdt) trips from the arrival-time distributions of their "coworkers". coworkers are defined as passengers who (likely) work at the same place, used to arrive to work at the same time, but live in other parts of the city and thus unaffected by the EBD. Specifically, we start with passengeres making (jdt) trips in 2016 (the main sample period) and trace them back to 2015 (the pre-EBD auxiliary period), to identify those who still traveled in the same OD pair j. For each such passenger i, we then select as coworkers those who traveled to the same destination station D during the morning rush hours in 2015 and arrived within the same 15-minute bin as i. These coworkers are then traced forward to 2016 and we retain those who continue commuting to station D during the morning rush hours. To ensure that coworkers are unaffected by the EBD, we exclude those whose trips originate from EBD stations or whose routes intersect with them. The optimal arrival time of each passenger i is measured as the median arrival time of his coworkers in 2016, and we aggregate these values to the (jdt) trip level to obtain (t^{OA}_{jdt}).

4.1.5 Passenger Income Distribution

Passengers may have heterogeneous preferences for trip characteristics. We are particularly interested in preference heterogeneity by income levels, which has implications for the distributional impacts of crowding and crowding-reducing policies. The smartcard data do not have information on user demographics. We match passenger income by matching outside data sources.

We use the grid-level commuting flows data provided by Baidu Maps, a leading provider of digital map services in China. Baidu Maps is a popular location service provider on smart devices and is embedded in many software applications. Baidu Maps is able to track millions of smart mobile devices (mostly smartphones) with precise and high-frequency location data. In particular, it can decide a device's usual nighttime location (which it refers to as home) and its usual daytime location (work). The commuting flows data used here contains the number of commuters between all pairs of home and work locations in Beijing. Each location is a small grid with an edge of about 700 meters (and an area of about 0.5 square kilometers). The commuting flows data are based on the usual location patterns in the three months ending in November 2021. There are 7.6 million unique devices with home locations distributed across some 15,000 grids.[8] Importantly, the data also reports the share of commuters in each home-work location pair in each of the five income bins. User income is inferred by Baidu Maps based on the information it has access to on the device using a machine learning algorithm. The information include the make and model of the device, applications

[8]Beijing has 22 million residents and a land area of 16,000 square kilometers.

installed on the device, some disclosed information to Baidu and its partnering applications, and of course, detailed location information. We do not have those underlying user characteristics, the exact algorithm Baidu uses, or the training data it uses to validate the imputations.

Three hurdles need to be bridged in order to map the income distribution of commuters to that of subway passengers in specific *OD* pairs. First, we need to convert the income distribution, which is from 2021, to the income distributions in around 2016. Second, the commuting flows data includes trips in all modes of transportation, we need to predict which travel using the subway. Third, we need to assign predicted subway passengers to *OD* pairs.

For income distribution conversion and mode choice prediction, we use the 2015 Beijing Household Travel Survey (BHTS). The 2015 BHTS was a large representative household survey (about 100,000 households and some 200,000 individuals). Aside from the usual basic demographic and socio-economic questions, the survey focuses on urban transportation. It records home and work (or school for students) locations, measured in Transportation Analysis Zones (TAZs). The TAZs are fine geographic areas that largely follows street and neighborhood boundaries. Beijing is divided into 1,911 TAZs, and the average size of a TAZ in the city center is about 1 square kilometer. The BHTS records the details of each respondent's travels in one day. For each trip during the day, the survey asks the departure time and TAZ, interchange TAZ(s) and arrival time, wait time at each interchange TAZ, modes of transportation for each leg of the trip, and the destination TAZ and arrival time. We assign the mode of the longest (in distance) leg as the main transportation mode of the trip.

We would like to know which commuters in the Baidu commuting flows data use the subway. To do so, we fit a model of mode choice using the BHTS with predictors that are also available in the Baidu data, i.e., income levels and origin-destination locations. The BHTS only has household income (in five bins), not individual income. We regress household income (converted to log RMB) on flexible functions of each working household member's gender, age, education level, industry and occupation, the interactions of these characteristics, and a household fixed effect. The estimated coefficients of the regression are used to predict individual income subject to the constraint that the sum of all working members' incomes is equal to the observed household income.

We then fit a multinomial choice model of mode choice, where each observation is a commuting trip (between home and work). The dependent variable is the main mode used for that trip. Modes of transportation include the subway, bus, car, with other modes (bicycles, electric scooters, walking) as the base category. Explanatory variables include log individual income, trip distance, and distances to the nearest subway station from the centroids of home and work TAZs. Log individual income and distances to the nearest subway stations are the only information available from the Baidu commuting flows data.

Table 4 reports the results from the multinomial logit estimation. Higher income is positively correlated with the use of all transportation modes other than walking, biking, or scooting. High-income individuals are more likely to use the subway or a car. Trip distance is also positively correlated with

Table 4: Estimation of Travel Mode Choice

	(1) Subway	(2) Bus	(3) Car
Income (1,000 RMB)	0.271	0.071	0.202
	(0.007)	(0.004)	(0.004)
Income^2	-0.006	-0.001	-0.003
	(0.000)	(0.000)	(0.000)
Trip Distance (km)	0.473	0.437	0.412
	(0.004)	(0.004)	(0.004)
Observations		80,259	
Pseudo R^2		0.258	
Chi^2		51456.3	

Notes: The estimates are based on the multinomial logit model:

$$P(y_i = j) = \frac{\exp(\alpha_j + \beta_{1j}\text{Income}_i + \beta_{2j}\text{Income}_i^2 + \gamma_j\text{Distance}_i)}{1 + \sum_{k\in\{\text{Sub, Bus, Car}\}} \exp(\alpha_k + \beta_{1k}\text{Income}_i + \beta_{2k}\text{Income}_i^2 + \gamma_k\text{Distance}_i)}$$

Standard errors in parentheses. The baseline reference group is “other” travel mode.

Table 5: Predicted Travel Mode Probabilities for Low and High Income Passengers

Group	Passenger Attributes: Income (1,000 RMB)	Passenger Attributes: Distance (km)	Predicted Probability: Subway	Predicted Probability: Bus	Predicted Probability: Car	Predicted Probability: Other
Low Income	4.95	22	0.323	0.390	0.286	0.0003
High Income	9.67	22	0.417	0.264	0.319	0.0001

Notes: Probabilities are calculated based on the multinomial logit estimates. "Low Income" represents a passenger with 4,950 RMB monthly income (35th percentile), and "High Income" represents 9,670 RMB (65th percentile). Trip distance is fixed at 22 km (average trip distance).

the use of the subway.

Table 5 illustrates the mode choices of a group of hypothetical low-income passengers with a monthly income of 4,950 RMB (35th in the income distribution) and a group of high-income passengers who make 9,670 RMB a month (65th in the income distribution) for a trip of 22 kilometers. The multinomial logit model predicts that 32.3% of the low-income people ride the subway, 39% of them ride the bus, and 29% drive. For high-income people, the share of taking the subway, the bus, and driving is 42%, 26%, and 32%, respectively. A 22 km trip is a long trip, the model predicts that essentially no one chooses to walk, bike, or scoot.

We then convert the aggregate income distribution from the Baidu commuting flows data to the imputed individual income distribution from the BJHT. First, we deflate income in the Baidu data (in 2021) into the 2015 RMB using chained consumer price indices. We then rescale the deflated income and fit a β-distribution that matches the median and standard error of the individual income distribution from the BHTS.

We convert the Baidu commuting flows data into individual-level data with information on each person's work and home locations and income. Feeding these characteristics into the fitted logit model. Individuals with more than 50% predicted probability are identified as subway passengers.[9] Appendix Figure B.1 shows the fitted distribution of subway users' incomes deflated to the 2014 RMB. After this step, we have each implied commuter from the Baidu data with its accurate home and work locations, converted income level, and imputed mode choice (riding the subway or not).

We assign imputed subway users from the Baidu data to *OD* station pairs. We define a station's catchment as an area within a three-kilometer radius. Data from the BHTS shows that very few people choose to ride subway if they live or work more than three kilometers away in straight-line distance from a subway station. Catchment areas from different stations can overlap. Users in multiple catchment areas are probabilistically assigned to the station with weights equal to the inverse of distance. For example, suppose a worker lives in the catchment areas of both stations S_1 and S_2, the distance from his home to S_1 is 1 km while that to S_2 is 0.5 km, then we assign the worker to S_1 with a probability of $1/3$ and to S_2 with a probability of $2/3$. Similarly, the worker's workplace can be probabilistically assigned to a station. Applying the same algorithm to all imputed subway users in the Baidu data, we define the user pool of a station pair *OD*. Notice that commuting trips are in both directions. The user pools of station pairs *OD* and *DO* are the same.

For each market (*OD* pair by date), 150 subway passengers are randomly drawn from the fitted income distribution of the *OD* pair's user pool. About a third of *OD*-dates do not have income data because insufficient number of predicted subway users. This is either due to few commuters in the *OD* pair in the original Baidu data, or due to parameters from the mode choice model predicts that few chooses the subway. Those markets are small. Collectively, they account for a small share of the total ridership in the smartcard data.

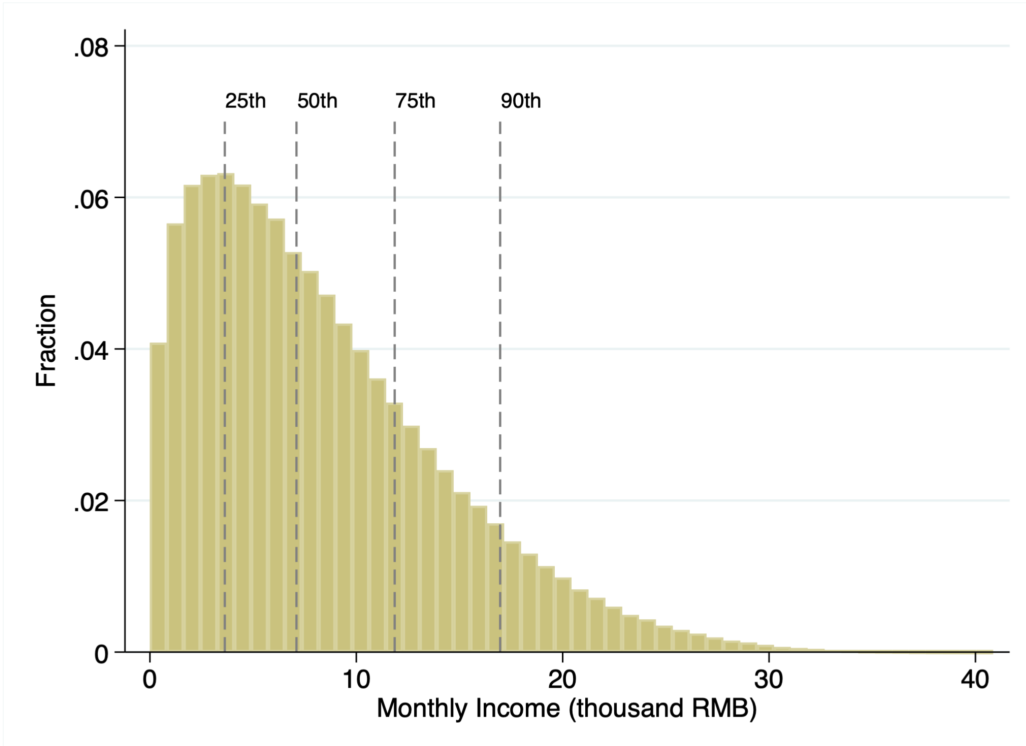
Figure 5 plots the distribution of simulated monthly income of imputed subway riders. The median monthly income of subway riders is about 8,000 RMB. 90^{th} percentile is at around 18,000 RMB.

5 Sample and Stylized Facts

This section introduces the sample selection and presents the summary statistics. We then present several stylized data patterns that help motivate our modeling and identification choices.

[9] Alternatively, we can calculate the probability of using subway for each TAZ pair and income bin using the BJHS, and assign the probability to the corresponding OD pair and income level in the Baidu data. However, the BJHS is too small to fill all location-and-demographic cells observed in the Baidu data. The parametric approach described above allows us to predict subway users for all implied individuals in the commuting flows data.

Figure 5: Histograms of the Simulated Monthly Income (thousand RMB)



Notes: The graph shows the simulated distribution of monthly income of potential subway riders.

5.1 Sample and Summary Statistics

Our sample includes trips originating from the 16 EBD stations and 16 control stations and end at any of the 342 stations in operation. We include trips with a tapping-in time between 6:30 and 8:30 AM. The cutoff for the EBD is at 7 AM by tapping-in time. Focusing on the time window around the time cutoff allows us to exploit exogenous price variation in the EBD stations. While the first train departs the terminal station around 5 AM in the four lines in the sample, the ridership is relatively low in the first hour or so (Figure 1), and we therefore start the sample at 6:30 AM.

The sample covers five non-continuous weeks that span across March, June, September, November, and December in 2016. Weekends and holidays are excluded as the EBD does not apply on those days. In addition, travel patterns and passenger compositions are substantially different between work and non-work days. This leaves us with a sample of 17 working days.

We define a *market* as an OD-by-date within the selected time window. In each market, passengers choose when to tap into the origin station. Each *product* is a 15-minute window. Therefore, there are eight products in each market. The outside good is traveling within the same market during the morning rush hours but using alternative modes of transportation. There are 4.9 million subway trips across 144,060 markets.

Many markets are small, not all products are chosen in some markets. We focus on markets that are big enough and have at least one trip in each of the eight products. This restriction excludes 72% of the markets and 22% of the trips.

Passengers choose when to leave according to *expected* crowding, which is constructed based on historical crowding levels. A further 26% (40,190 out of 822,437) markets, including 22% (822,437 out of 3,783,770) of the remaining trips are excluded from the sample.

The final sample includes 2.96 million trips from 29,920 markets. There are 13,130 markets from the EBD stations, including 1.23 million trips in 1,435 unique OD pairs. There are 16,790 markets in the control sample, including 1.74 million trips in 1,736 unique OD pairs.

Table 6: Summary Statistics of Trips from Control Stations

Variables	(1) Treatment: EBD Policy Covered OD pairs					(2) Control: EBD Policy Free OD pairs				
	Mean	SD	P10	P90	N	Mean	SD	P10	P90	N
# of OD pairs	1,304	0	1,304	1,304	94,616	1,438	0	1,438	1,438	107,952
# of days	17	0	17	17	94,616	17	0	17	17	107,952
# of tap-in time intervals	8	0	8	8	94,616	8	0	8	8	107,952
log_share	-3.62	1.11	-5.08	-2	94,616	-3.55	0.93	-4.79	-3	107,952
log_share_frequent	-3.61	1.12	-5.09	-2	93,710	-3.55	0.94	-4.80	-2	107,066
# of passengers by subway	11.36	22.35	2.00	22	94,616	12.33	16.19	2.00	27	107,952
# of frequent passengers by subway	10.71	21.65	1.00	21	94,616	11.58	15.61	2.00	26	107,952
# of passengers by other modes	681.09	2579.43	26.00	1,462	94,616	406.19	798.23	35.00	1,055	107,952
# of frequent passengers by other modes	616.19	2214.69	23.00	1,371	94,616	380.17	749.57	31.00	990	107,952
dist (km)	23.14	8.47	10.96	34	94,616	21.09	7.47	11.81	31	107,952
total travel time (minute)	54.47	19.31	26.02	78	94,616	45.85	15.47	26.18	66	107,952
in-train travel time (minute)	37.99	14.45	18.00	56	94,616	35.68	13.53	18.00	53	107,952
price (RMB)	5.14	1.13	3.50	7	94,616	5.37	0.82	4.00	6	107,952
rescheduling: abs of deviation from optimal arrival time (minute)	15.95	12.18	2.60	33	93,592	14.20	10.52	2.41	28	107,032
crowding (person minute/m^2)	129.22	65.84	41.99	219	94,616	121.02	52.63	50.77	187	107,952
historic crowding (person minute/m^2)	131.13	66.25	43.11	222	94,616	122.99	52.98	52.31	190	107,952
crowding per minute (person/m^2)	3.36	1.36	1.66	5	94,616	3.53	1.32	1.77	5	107,952
IV: # of ACP: infrequent riders (person/m^2)	0.09	0.06	0.03	0	94,616	0.10	0.07	0.04	0	107,952
IV: # of ACP: total riders (person/m^2)	0.93	0.76	0.24	2	94,616	1.15	0.95	0.33	2	107,952
IV: # of ACT: infrequent riders (person minute/m^2)	1.67	1.18	0.54	3	94,616	1.65	1.07	0.58	3	107,952
IV: # of ACT: total riders (person minute/m^2)	16.36	13.55	4.87	34	94,616	17.93	13.92	5.45	34	107,952
Residual of # of infrequent ACT	0.00	0.34	-0.32	0	94,616	0.00	0.36	-0.38	0	107,952
Residual of # of total ACT	0.00	5.61	-5.81	6	94,616	0.00	5.72	-5.23	6	107,952

Notes: R^2 in the two residual regressions are 0.9 and 0.83 for infrequent and total ACT, respectively. The associated specification is $ACT_{jdt} = \kappa_{jd} + \lambda_t + \xi_{jdt}$.

Table 6 reports summary statistics of the key variables for the treated and control groups. The average number of passengers in a product (a 15-minute time bin on a given day in an OD pair) is 11.6 in the treated sample and 13.0 in the control sample. Trips from both EBD and control stations are long. The average travel distance is 23.7 km for the EBD stations and 21.7 km for the control stations. The average travel time for trips originating from the EBD stations is 55.7 minutes, with the 10th percentile at 26.8 minutes and the 90th percentile at 80 minutes, compared to an average of 47.3 minutes for the control stations. The average price paid for the trip, after discounts, is 5.19 RMB in the EBD sample and 5.43 RMB in the control sample. Finally, the average deviation from the optimal arrival time is 16.1 minutes for the EBD stations and and 14.5 minutes for the control stations.

Passengers in the sample experience high levels of crowding. For trips originating from the EBD stations, the average in-train crowdedness is 3.3 persons per square meter. For the top 10% trips with the highest crowding experience, there are about five passengers in one square meter. We also measure the total experienced crowding. Its average, at 130 person-minutes per square meter in the EBD stations, is the product of multiplying the average crowding experienced during the trip (3.31 persons per square meter) by the average time on a train (39.3 minutes). The total experienced crowding captures the total discomfort due to crowding, which is the product of the intensity of the pain and the length of the suffering. The experienced total and average crowdedness is similar for trips originating from the control stations.

The crowding caused by accidental companion passengers (ACPs) serves as the instrumental variable for crowding. For trips originating from the EBD stations, ACPs contribute an average of 0.84 persons per square meter, accounting for about 25% of the measured crowding. Infrequent subway users represent a small share of overall ridership, contributing 0.08 persons per square meter. When weighted by overlapping travel time, the accidental companion time (ACT) of total riders amounts to 14.7 persons per square meter, while that of infrequent riders is 1.5 persons per square meter. Comparable figures are observed for trips originating from the control stations. The relevance of the instrument is testable. The last two rows show that there is abundant variation in the IV after controlling for market and time-bin fixed effects. The standard deviation of crowding caused by infrequent ACT is 1.23, its residualized version has a standard deviation of 0.33.

5.2 Stylized Facts

This subsection describes several stylized facts relevant to the identification and interpretation of our main results, which are described in the next section.

5.2.1 Crowding Does Not Cause Delay

Excessive time cost is a potential confounding factor for crowding. For example, when buses are crowded, the road is usually also congested, so the disutility from crowding is confounded by that from longer travel time. A unique feature of subway trips is that crowding in the train does not lead to longer travel times. Thus this is an ideal setting to identify disutility from crowding as a pure disamenity.

To test whether crowding is associated with delay, consider the following regression equation:

$$\bar{T}_{jdt} = \beta_1 Crowd_{jdt}^{\text{in-train}} + \beta_2 Crowd_{jdt}^{\text{in-station}} + \kappa_{jd} + \lambda_{lt} + \xi_{jdt}. \tag{13}$$

Each observation is a product (identified by an OD pair, a date, and a 15-minute time bin), and we use

Table 7: The Effects of Crowding on Average Travel Time

Quantile of travel time	p10	p25	p50	p75	p90	p99
	(1)	(2)	(3)	(4)	(5)	(6)
in-train crowding per minute	0.2332	0.2491	0.2704	0.3000	0.3341	0.4602
	(0.0118)	(0.0138)	(0.0176)	(0.0239)	(0.0318)	(0.0626)
in-station ridership per minute	0.0031	0.0033	0.0034	0.0037	0.0040	0.0052
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0002)
quantile of travel time	33.26	34.51	36.18	38.48	41.13	50.81
mean crowding per minute (persons/m^2)	2.87	2.87	2.87	2.87	2.87	2.87
mean in-station ridership per minute (poersons)	667.72	667.72	667.72	667.72	667.72	667.72
OD-Day FE	X	X	X	X	X	X
Line-Time FE	X	X	X	X	X	X

Notes: This table reports the quantile regression that shows the relationship between crowding and travel delays. We distinguish between in-train crowding and in-station crowding. There are 33.4 million observations in total.

the baseline sample. $\bar{T}_{jdt}$ is the average travel time, measured in minutes, for trips choosing product jdt. We separate two sources of crowding: in-train and in-station. β_1 captures lower train speed due to carrying more passengers; β_2 may reflect slower walking speed due to crowdedness in the station, longer waiting time for the elevator, or in some extreme cases, failing to board a train. κ_{jd} captures the market-specific shocks that could affect the travel time cost, λ_{lt} is the line-by-time fixed effect, which captures train frequency, which varies by line and potentially by the time of the day. We estimate Equation (13) using the OLS.

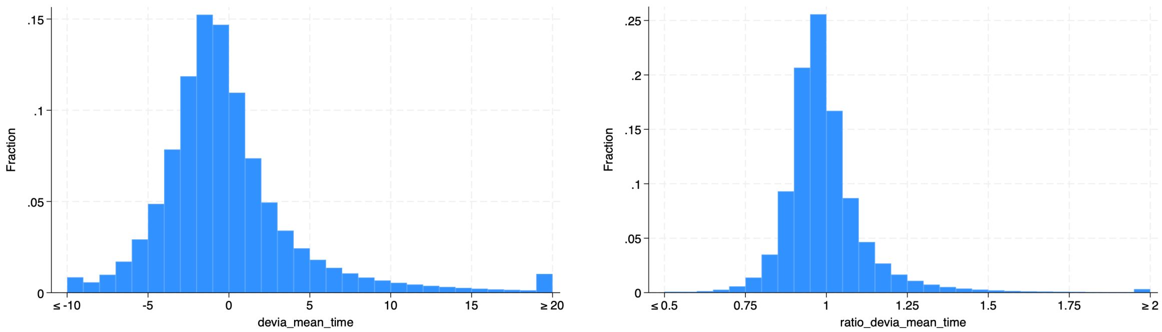
Table 7 reports the effects of in-train and in-station crowding on travel time across quantiles of the travel time distribution. The results show that increasing in-train crowding by 1 person/sqm slows the entire trip by 0.23 to 0.46 minutes, depending on the quantile. Given that average travel times range from 33 to 51 minutes, this effect represents less than a 1% increase in total travel time. Moving from the 10th percentile of in-train crowding (1.69 persons/sqm) to the 90th percentile (4.97 persons/sqm) slows down an average trip (36 minutes in train) by less than 1 minute, or about 2.5% of the total travel time. To put the magnitude in perspective, the median full-time worker makes 8,000 RMB per month, or RMB 50 per hour. Assuming the value of time on the subway is half the hourly wage, a 1-minute delay costs RMB 0.4. This is about 8% of the mean fare (5.19 RMB), or 3% of the total value of the in-train time cost for the trip (15 RMB).

Higher levels of in-station crowding also lengthen travel time, although it is not the focus of the paper. Having an additional 100 passengers in the same station, either during embarking, transferring, or exiting, constituting an 15.0% increase from the mean, leads to a less than 0.4-minute delay in the average trip, or about 0.9% increase from the average trip time cost.

5.2.2 Crowding Does not Lead to Rerouting

The measure of crowding depends on the crucial assumption that all passengers take the optimal route, which is the route that costs the least amount of expected time, which in turn is based on train

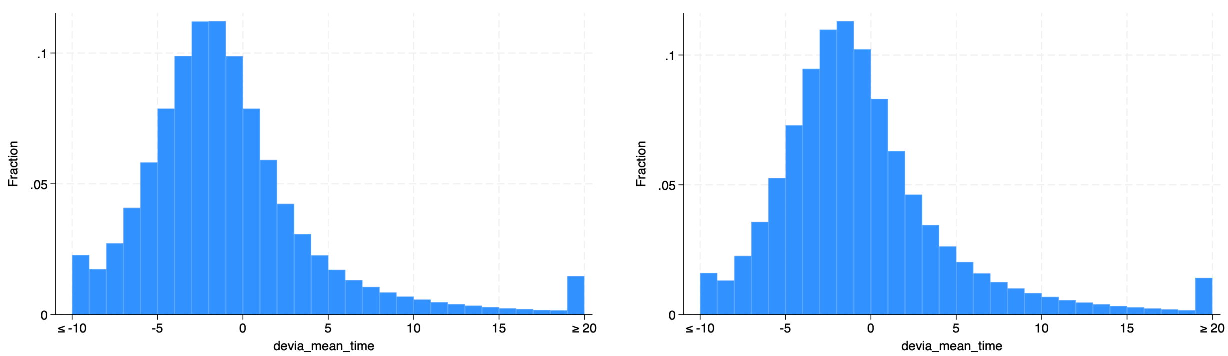
Figure 6: Histograms of Deviation from the Mean Travel Time



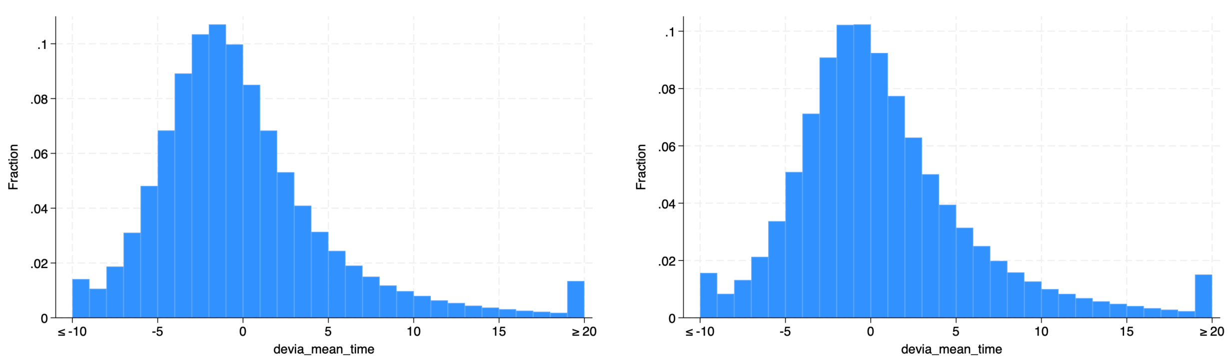
(a) Minute (b) Ratio

Notes: The graphs show the residualized distribution (from OD pair fixed effects) of travel time deviation (panel a) and ratio deviation (divided by mean travel time in the OD, panel b).

Figure 7: Deviation from the Mean Travel Time: by Quartiles of Crowding



Panel A: Least crowded quartile Panel B: Second quartile

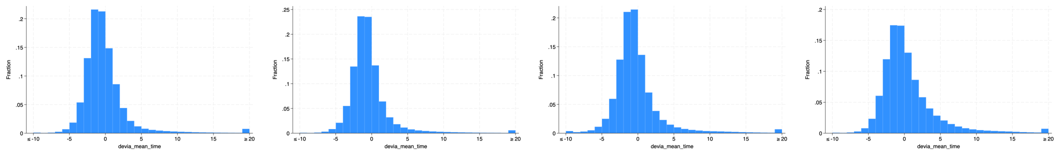


Panel C: Third quartile Panel D: Most crowded quartile

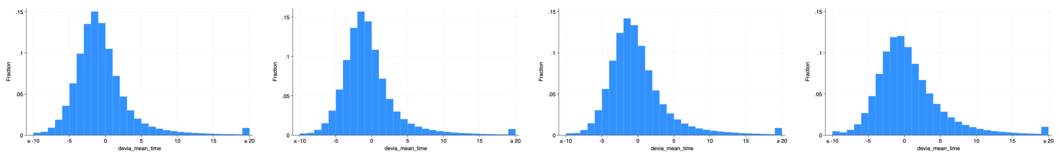
Notes: The graphs show the residualized distributions (from OD pair fixed effects) of travel time deviation by quartiles of crowding density.

Figure 8: Distribution of Travel Delays by Number of Transfers

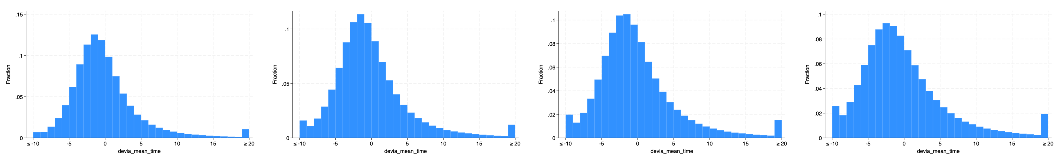
Panel A: Trips that Require No Transfer



Panel B: Trips that Require One Transfer



Panel C: Trips that Require Two or More Transfers



Notes: The graphs show the residualized distribution (from OD pair fixed effects) of travel time deviation. Panel A includes trips that require no transfer; Panel B includes trips that require one transfer; Panel C includes trips that require two transfers. In each panel, crowdedness in the lowest quartile is on the left, that in the highest quartile is on the right.

travel time and imputed accessing, exiting, and transferring times at each station.

A natural concern is whether some passengers choose alternative routes. An alternative route will inevitably increase the total travel time according to the time-cost parameters estimated in the previous steps. According to our calculations, the second-best route for almost all OD pairs costs substantially more time and is not associated with noticeably less crowding. While in reality, some passengers may deviate from the optimal route due to idiosyncratic reasons, those deviations lead to the classical error in the measurement of crowding and an attenuation bias. The instrumental variable can address such bias as long as the idiosyncratic deviations among infrequent ACPs are not correlated with those among the frequent users.

A more concerning scenario is where the choice of an alternative route is an endogenous response to crowding. As the optimal route gets more crowded, passengers are more likely to choose an alternative, less crowded route although it may take much longer. In this case, our measure overstates the actual experienced crowding, and the gap increases with the crowding value. Whether our estimate over- or under-states the true WTP depends on the characteristics of the alternative route (how much longer it takes and how much less crowded it is) and the value of time, which we do not model.

We argue that rerouting is sporadic and is unlikely to affect our estimates. As shown in Table 7, crowding does not significantly increases the *average* travel time, ruling out a substantial share of passengers taking alternative routes, which would generally be much more time-consuming. Figure 7 plots the distribution of observed travel time by quartiles of in-train crowding, where the median time cost of trips in the same OD-date-time bin is standardized to one. If crowding leads passengers to reroute, we should expect the distribution to have a fatter tail as crowding increases. The four graphs do not show such a pattern. Running the Kolmogorov-Smirnov test to detect differences between distributions, we cannot reject the four distributions are statistically identical.

Figure 8 further shows the distribution of travel time deviations by the number of transfers required. First, across all four graphs in each row, we find no evidence that in-train crowding induces passengers to reroute, as the distribution shapes do not exhibit fatter tails as crowding increases, regardless of transfer status. What does vary substantially is the extent of delay: trips without transfers exhibit the tightest distribution with minimal dispersion, while trips involving one transfer display noticeably wider tails, and those with two transfers show the largest delays. This pattern indicates that delay is mostly attributable to transferes rather than in-train crowding.

6 Estimation Results

6.1 Results from the Standard Logit Model

OLS Estimates: Table 8 reports the results from estimating Equation (4) using OLS. Note that after controlling for OD-day fixed effects and time fixed effects, price in levels is identified by trips from the EBD stations before and after the event time, and comparing trips from the control stations. Price is not identified by the functional form – if we use log price, the coefficient is still identified. Price is identified by the arbitrary cutoff of the EBD and the comparisons between EBD and control stations. The price coefficient is negative, the expected sign, without instrumenting.

We estimate a moderate cost of rescheduling. Deviating away from the ideal arrival time by one minute costs 0.16 RMB for all passengers. Deviating by 15 minutes (the sample mean) costs 2.40 RMB, or 4.8% of the average hourly wage (RMB 50). Deviating by one hour amounts to 19% of the average hourly wage. The estimated rescheduling costs are similar for frequent passengers.

Expected total crowding (persons per square meter times in-train travel length in minutes) is used as the main explanatory variable. Without accounting for the endogeneity of crowding, the coefficient associated with crowding is positive. This highlights the positive correlation between crowding and unobserved demand shocks.

Table 8: Standard Logit: OLS Estimation Results

	All passengers		Frequent passengers	
	(1)	(2)	(3)	(4)
price	-0.1545	-0.1533	-0.1551	-0.1537
	(0.0138)	(0.0138)	(0.0145)	(0.0145)
rescheduling	-0.0250	-0.0251	-0.0265	-0.0266
	(0.0005)	(0.0005)	(0.0006)	(0.0006)
crowding	0.0022	0.0023	0.0023	0.0024
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	200,624	200,624	187,928	187,928
R^2	0.7336	0.7343	0.7305	0.7312
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X
Mean rescheduling	15.02	15.02	14.90	14.90
Total crowding	127.4	127.4	127.4	127.4
MWTP to reduce rescheduling	0.162	0.164	0.171	0.173
by 1 min	(0.0153)	(0.0155)	(0.0167)	(0.0171)
MWTP to reduce crowding	-0.014	-0.0148	-0.0145	-0.0155
by 1 person·min/m^2	(0.0022)	(0.0022)	(0.00235)	(0.00247)

Notes: This table reports OLS estimates of Equation 4.

Instrument for Crowding and GMM Estimates: To address the endogeneity concern of crowding, we exploit the plausibly exogenous variation generated by infrequent accidental companion passengers (ACPs) and construct the instrumental variable accidental companion time (ACT) as the weighted sum of ACPs, with weights given by the shared travel time. Table 9 reports the first stage estimation result for crowding.[10] The estimates indicate a strong positive correlation between ACT and crowding, with a first stage F-statistic exceeding 650 for both the overall and frequent passengers. For example, the result in column (4) implies that a one-unit increase in ACT predicts an additional 10.34 person-minutes per square meter of crowding for frequent passengers.

Table 10 reports the results from estimating Equation (4) using GMM, where crowding is instrumented by infrequent ACT. Columns (1) and (2) report the results for all passengers, and columns (3) and (4) for frequent passengers. The coefficient on price is similar to that in the OLS. The marginal WTP for rescheduling is also similar to OLS, amounting to 2 cents per minute. The crowding coefficient is now negative, as expected. The MWTP for crowding reduction is around 0.6 cents per

[10] First stage results for additional instruments are presented in Appendix Table **??**

Table 9: First Stage Estimation for Crowding

	Crowding (K)			
	Sample: overall		Sample: frequent riders	
	(1)	(2)	(3)	(4)
price	-4.8031	-4.9164	-4.9401	-5.0644
	(0.5200)	(0.5191)	(0.5379)	(0.5366)
rescheduling	-0.3192	-0.2987	-0.3358	-0.3136
	(0.0202)	(0.0204)	(0.0210)	(0.0212)
ACT	10.2326	10.0174	10.5711	10.3379
	(0.3714)	(0.3897)	(0.3863)	(0.4052)
First Stage F-Test	759.1	660.8	748.9	651.1
Observations	200,624	200,624	187,928	187,928
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X

Notes: This table reports the first stage of the standard logit model in Equation 4.

person-minute/sqm. Consider a typical trip with average crowding of 3.31 persons/sqm and lasts 40 minutes in train, the WTP to reduce one person per sqm during the course of the subway ride for frequent passengers is 2.416 RMB (0.0604*1*40). Considering that the user pays about RMB 5.2 for the trip, the MWTP for crowding avoidance is substantial. Its order of magnitude is comparable to the rescheduling cost, which amounts to 3 RMB for a 15-minute deviation.

IV Validity: As discussed in Section 3.2.3, crowding is potentially endogenous because it is correlated with unobserved demand. To break this correlation, our idea is to exploit variation in crowding caused by other passengers whose travel decisions are plausibly uncorrelated with the unobserved demand driving bias in the OLS estimates. The instrument incorporates two features. First, we define ACPs as those who travel from an origin that is in the same line as O but in the upper stream, and travel to a destination that is sufficiently far away from D, making them unlikely to respond to the same demand shocks as the focal passenger. Second, we use only ACPs who are infrequent users, who exhibit different travel patterns from most of the sample, which is from frequent users.

We assess robustness by examining how results change when varying these two dimensions of the instrument. Specifically, we estimate three sets of models: (1) OLS, (2) GMM estimation where crowding is instrumented by ACT constructed from all ACPs, and (3) GMM with ACT constructed from infrequent ACPs. For each model, we report the estimated WTP for crowding reduction with varying exclusion range around destination station D and progressively tighten the exclusion around origin station O. Figure 9 presents three panels corresponding to different exclusion rules for ACPs near the origin station O: (i) no exclusion, (ii) exclusion of ACPs who travel from origin stations within 1 km of station O, and (iii) exclusion of ACPs from origin stations within 2 km of station O.

Table 10: Standard Logit: GMM Estimation Results

	All passengers		Frequent passengers	
	(1)	(2)	(3)	(4)
price	-0.2181	-0.2223	-0.2227	-0.2270
	(0.0162)	(0.0165)	(0.0171)	(0.0174)
rescheduling	-0.0284	-0.0283	-0.0301	-0.0300
	(0.0007)	(0.0007)	(0.0007)	(0.0007)
crowding	-0.0117	-0.0123	-0.0119	-0.0125
	(0.0011)	(0.0012)	(0.0012)	(0.0013)
First Stage F-Test	759.1	660.8	748.9	651.1
Observations	200,624	200,624	187,928	187,928
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X
Mean rescheduling (mins)	15.02	15.02	14.90	14.90
Mean crowding (persons·min/m^2)	127.4	127.4	127.4	127.4
MWTP for rescheduling avoidence	0.130	0.127	0.135	0.132
(RMB, per min)	(0.0100)	(0.0098)	(0.0107)	(0.0105)
MWTP for crowding reduction	0.0534	0.0553	0.0536	0.0552
(RMB, per person·min/m^2)	(0.0056)	(0.0058)	(0.0057)	(0.0059)

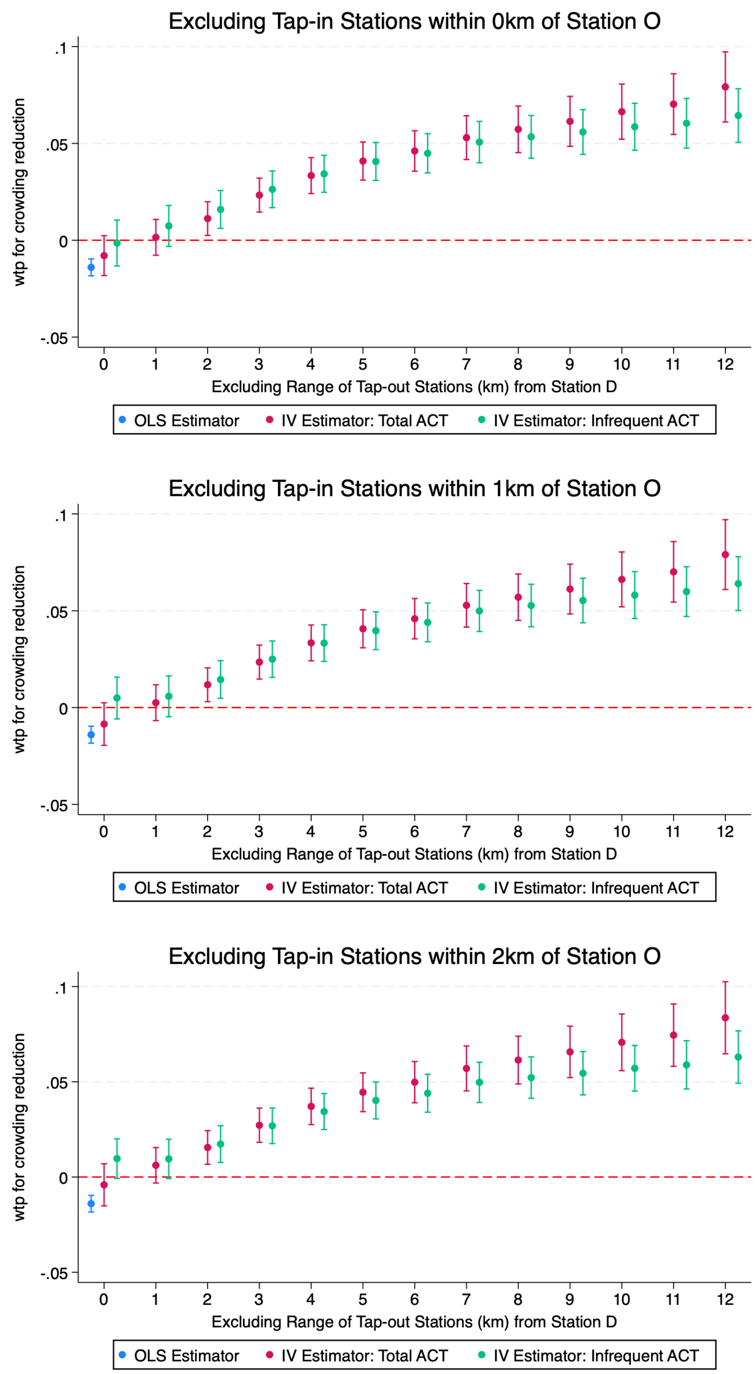
Notes: This table reports the GMM estimations of the standard logit model in Equation 4.

Within each panel, the horizontal axis varies the exclusion range for destination stations from 0 to 12 km.

Across all panels, the IV estimates based on ACT — whether constructed from all ACPs or restricted to infrequent ACPs — are non-negative and increase steadily as the exclusion range for destination stations expands. The estimates stabilize around 0.05–0.07 once stations beyond roughly 8 km are excluded, which justifies our choice of 8 km as the baseline cutoff. Figure B.3 shows that 8-km radius captures a small fraction of Beijing subway network, and Figure 10 indicates that more than 80% of ACPs travel to D's more than 8 km away from D. The patterns remain robust when ACPs from origins within 1 km or 2 km of station O are also excluded. These results suggest that the endogeneity of crowding mainly comes from common shocks at the destination rather than the origin. Furthermore, the similar estimates based on ACT from infrequent or all users indicate that the endogeneity is not due to general correlated shocks among frequent users.

Non-linear WTP: We allow for a non-linear effect of crowding by including a quadratic term in equation (4) and estimate the model using the additional instruments introduced in Section 3.2.3. The results for frequent passengers, reported in Table 11, show that while the linear term of crowding is small and imprecisely estimated, the quadratic term is negative and significant, indicating increasing marginal disutility from higher levels of crowding. The distribution of MWTP is presented in the bottom panel: at the 25th percentile, MWTP for crowding reduction is around 0.1 cents per person-

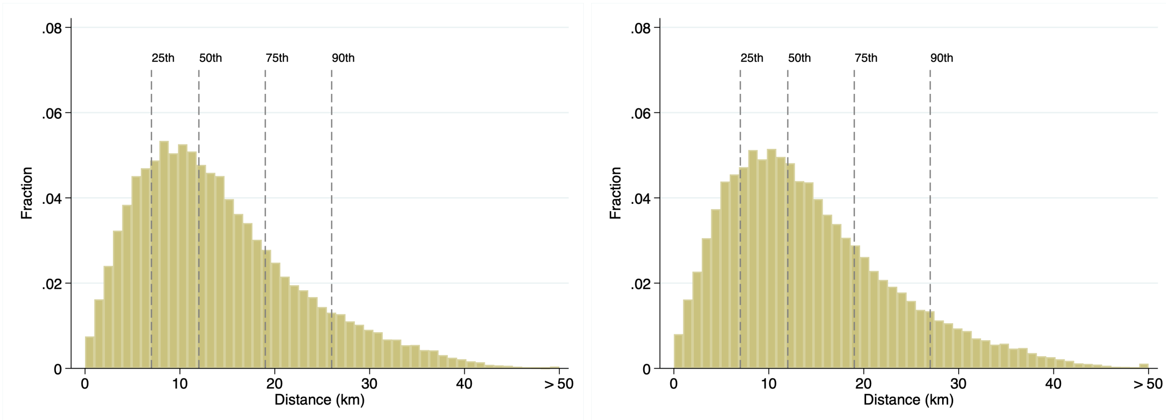
Figure 9: Source of Crowding Endogeneity



Notes: The graphs report MWTP estimates of reducing crowding by one person per square meter for one minute (one unit of k). The various estimates include OLS and 2SLS, using ACT from all riders or infrequent riders only, and vary by the size of the neighboring origin and destination stations to exclude.

minute/sqm, rising to about 0.6 cents at the 95th percentile. Figure 11 illustrates this relationship more directly, showing that MWTP increases monotonically as crowding rises.

Figure 10: The Histogram of Distance from ACPs' Tap-out Stations to Station D



(a) ACP: all riders (b) ACP: infrequent riders

Notes: The graphs show the density distribution of distances between D and ACP's destination D'. Cumulative percentiles are marked in dashed verticle lines. In Panel (a), all subway riders can be an ACP; in panel (b), only infrequent riders can serve as an ACP.

6.2 Results from the Random Coefficient Logit Model

In this section, we relax the homogeneous preference assumption and allow passengers' marginal utility from price, crowding and rescheduling to vary by income. We use four additional instruments introduced in section 3.2.3 - interactions between an indicator for the EBD stations and an indicator for the tap-in time before 7, and linear, quadratic, and cubic time trends - to estimate the random coefficient model, following the procedure discussed in section 3.1.2.

We report the estimation results of the random coefficient model and the associated MWTP for crowding reduction and rescheduling avoidance in Table 12 and Table 13, respectively. There are several key findings for heterogeneity in preference parameters. First, the median MWTP for crowding and rescheduling is 0.068 RMB and 0.210 RMB, respectively, which are close the estimates from the standard logit model presented in Table 10. However, the mean MWTP is larger than the standard logit estimates, particularly for crowding. Second, the positive and significant coefficient α_1 indicates that passengers with a higher income are less sensitive to the price. On the other hand, the negative and significant coefficients β_1 and ρ_1 imply that higher-income passengers experience larger disutility from both crowding and rescheduling. Compared with a passenger with an average income, the marginal disutility from crowding is more than three times larger for a passenger at the 75th percentile of the income distribution, and the marginal disutility from rescheduling is more than twice as large.

Figure 12 further shows the fitted line of the MWTP estimate over income levels with 95% confidence intervals. The MWTP for both crowding and rescheduling increases with income. The gradient in WTP is driven by two complementary forces: higher-income passengers derive larger disutility

Table 11: Standard Logit: GMM Estimation Results with Non-linear WTP

	Sample: frequent riders	
	(1)	(2)
price	-0.138915	-0.128100
	(0.018432)	(0.018414)
rescheduling	-0.027779	-0.027254
	(0.000893)	(0.000874)
crowding	-0.000044	0.002913
	(0.003772)	(0.003792)
crowding^2	-0.000017	-0.000024
	(0.000010)	(0.000009)
First Stage F-Test	81.81	83.23
OD-Day FE	X	X
Time FE	X	
Time-Day FE		X
Mean rescheduling	14.90	14.90
Mean crowding	127.4	127.4
MWTP for crowding reduction: p25	0.0209	0.0088
	(0.0140)	(0.0167)
p50	0.0317	0.0252
	(0.0082)	(0.0096)
p75	0.0411	0.0396
	(0.0063)	(0.0068)
p90	0.0511	0.0549
	(0.0099)	(0.0109)
p95	0.0574	0.0645
	(0.0134)	(0.0151)

Notes: This table reports the GMM estimations of the standard logit model with a quadratic term in crowding density.

from crowding and rescheduling while being less price sensitive, resulting in a steeper MWTP to avoid them.

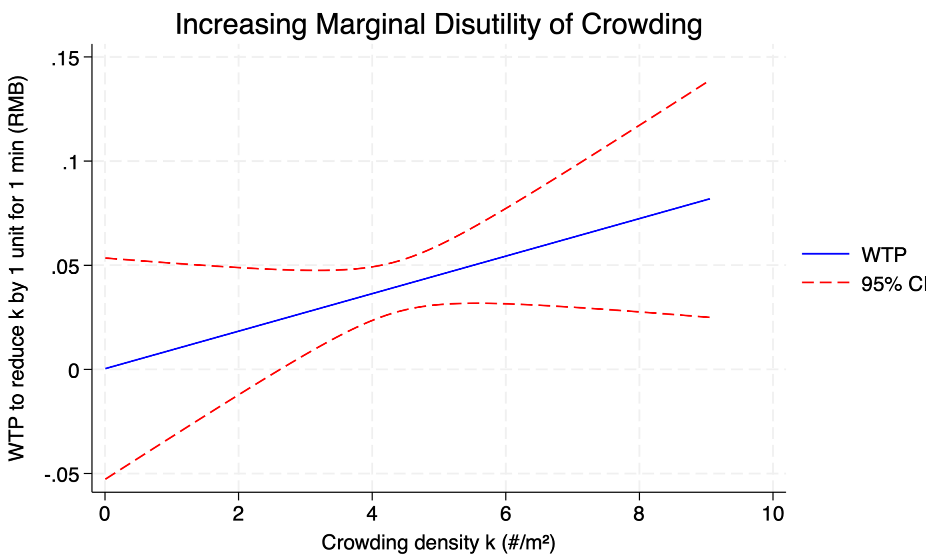
7 Welfare and Counterfactuals

7.1 Setup

To fully capture the crowding-induced externality and reduce the computation burden, we consider a one-way segment of subway, where passengers travel from O to D. There are two income groups $g \in l, h$, with $POP_l = POP_h = \frac{1}{2}POP$ where POP denotes total population in the economy. The utility of taking the subway is given as:

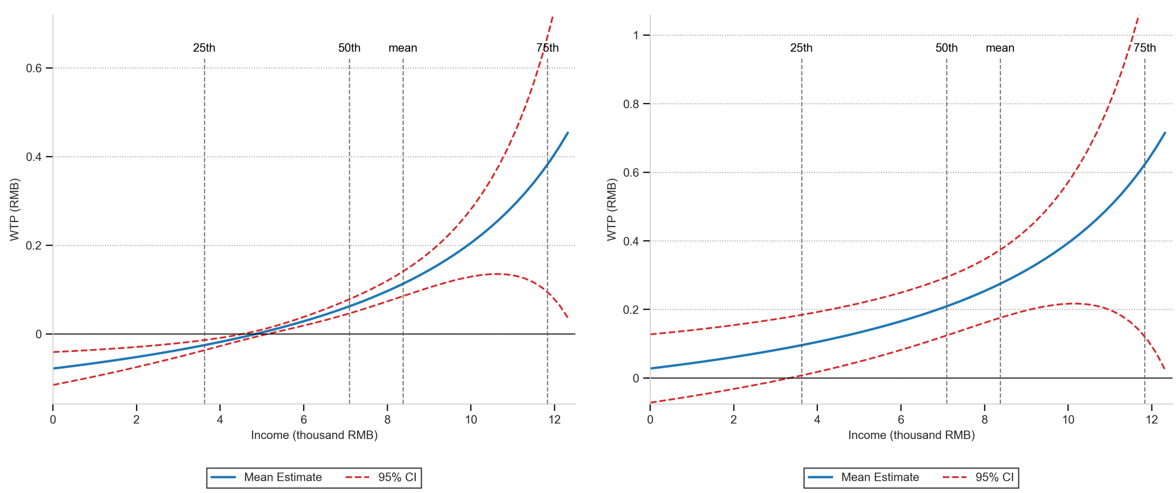
$$U_g = \underbrace{A + \alpha_g P + \beta_g \text{Crowd}}_{V_g} + \varepsilon \tag{14}$$

Figure 11: Increasing Marginal Disutility of Crowding



Notes: The graph shows the marginal WTP to reduce crowding density by 1 unit ($person/m^2$) as a function of crowding density. Dashed lines indicate the 95% confidence interval of the estimate.

Figure 12: MWTP for Crowding and Recheduling by Income



(a) WTP to Reduce Crowding by 1 person/m^2 · minute (b) WTP to Reduce Rescheduling by 1 minute

Notes: Panel (a) shows the marginal WTP to reduce crowding density by one unit ($person/m^2$) as a function of the passenger's income. Panel (b) shows the marginal WTP to reduce deviation from the optimal arrival time by one minute. Dashed lines indicate the 95% confidence interval of the estimate. Verticle dashed lines show the selected percentiles in the income distribution.

$$= A + \alpha_g P + \beta_g \underbrace{\frac{\text{TravelTime}}{\text{CarArea} \times M}}_{L} N + \varepsilon$$

Table 12: Random Coefficient Logit Estimation Results

	(1)
price:	
Mean coefficient (α_0)	-0.1246
	(0.0148)
Interaction passenger income (α_1)	0.0154
	(0.0069)
crowding:	
Mean coefficient (β_0)	-0.0156
	(0.0004)
Interaction passenger income (β_1)	-0.0046
	(0.0001)
rescheduling:	
Mean coefficient (ρ_0)	-0.0343
	(0.0041)
Interaction passenger income (ρ_1)	-0.0032
	(0.0013)
Objective Value	2.14
OD-Day FE	X
Time FE	X
Mean monthly income (1,000 RMB)	7.84

Notes: This table reports the results from the random coefficient model described in Section 3.1.2.

where A is a common mean utility, P is the fare paid by passengers, N is the number of subway passengers, M is the number of cars in a train, and ε is the idiosyncratic shock following a type-I extreme value distribution. We use V_g to denote the deterministic part of the utility and use L to denote the congestion exposure term with the unit of minutes per square meter. Parameters satisfy $\alpha_g < 0$ (price sensitivity) and typically $\beta_g < 0$ (crowding disutility).

For a two-choice logit model between subway and surface transportation (outside good), the probability and market share of taking the subway for group g are $s_g(V_g) = \frac{e^{V_g}}{1+e^{V_g}}$ and $N_g = POP_g \cdot s_g(V_g)$, respectively, and the total number of subway passengers is given as $N = \sum_g N_g$.

7.2 Welfare Analysis Framework

The welfare analysis builds on the distinction between marginal private demand (PD) and marginal social demand (MSD) for subway ridership, as illustrated in Figure 13. The vertical distance between PD and MSD reflects the marginal external cost of crowding. The deadweight loss associated with unpriced crowding is therefore captured by the area between private and social demand at the market equilibrium.

Table 13: MWTP for Crowding and Rescheduling by Income

	(1)
MWTP for crowding reduction:	
10th percentile	-0.0604
25th percentile	-0.0256
50th percentile	0.0684
mean	0.1252
75th percentile	0.4028
90th percentile	33.439
MWTP for rescheduling avoidance:	
10th percentile	0.0638
25th percentile	0.1034
50th percentile	0.2102
mean	0.2749
75th percentile	0.5906
90th percentile	38.1704

Notes: This table reports the marginal WTP for crowding reduction and rescheduling avoidance at various percentiles of the income distribution, with estimates from the random coefficient model.

To calculate the welfare impacts, we first derive the inverse private demand curve $PD(N)$. Fix any candidate aggregate ridership N, the price P that rationalizes N as an equilibrium must satisfy the adding-up condition:

$$N \; = \sum_{g\in\{\ell,h\}} POP_g \, s_g\big(A + \alpha_g P + \beta_g LN\big).$$

Define $f(P;N) = \sum_g POP_g s_g(A + \alpha_g P + \beta_g LN) - N$. For each N, we solve for P such that $f(P;N) = 0$ by bracketing and bisection. This implicit mapping is the inverse private demand $PD(N)$.

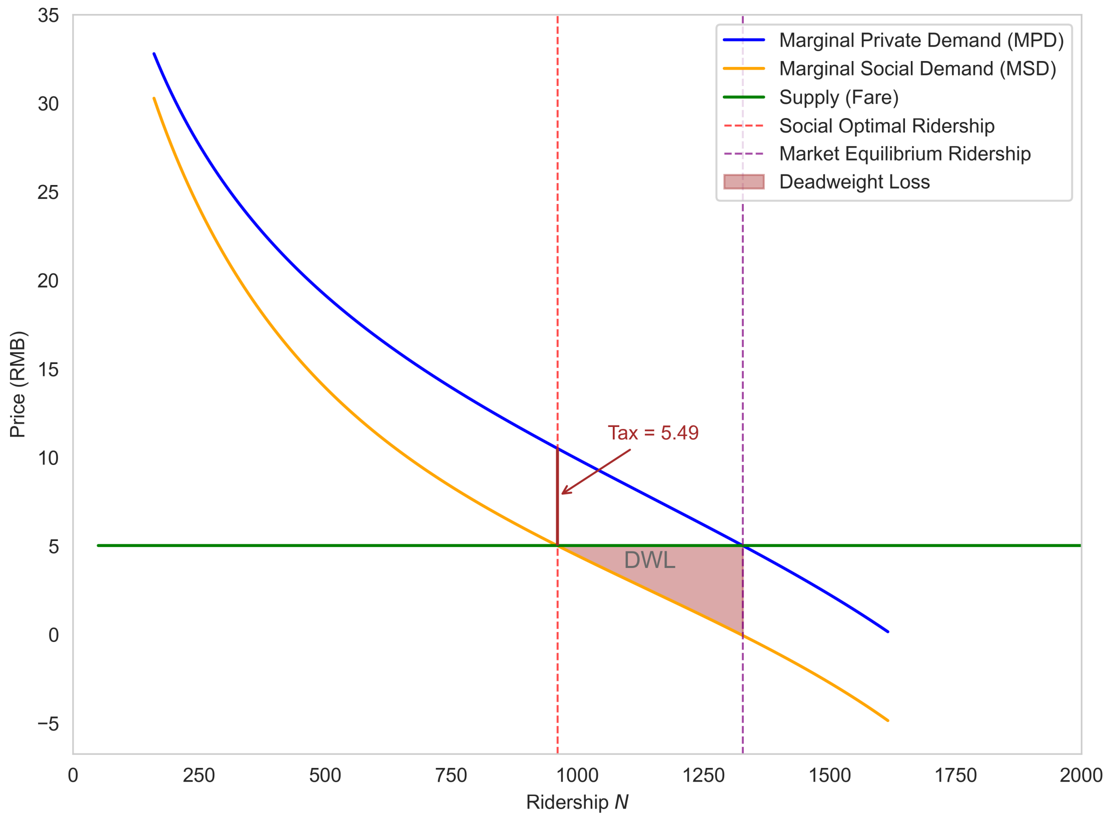
Let $N_g(P,N) = POP_g s_g(A + \alpha_g P + \beta_g LN)$ be the group-specific subway ridership. The marginal social demand (MSD) equals the price at which a passenger's MWTP for crowding is equal to the marginal external cost of the crowding they impose on others. Based on the logit-with-crowding setup built in previous section, it is convenient to work with the *externality wedge*:

$$w(N) \;\equiv\; PD(N) - MSD(N) \;=\; L \sum_{g\in\{\ell,h\}} \frac{\beta_g}{\alpha_g} N_g\big(PD(N), N\big).$$

The intuition is that each additional passenger raises crowding linearly in L, and the monetized crowding marginal effect for group g is (β_g/α_g) because price is the numeraire.

Model parameters are calibrated based on the data and the empirical estimation results reported in Section 6. Specifically, we set $\alpha\_\ell = -0.1738$, $\beta\_\ell = -0.0007$, $\alpha\_h = -0.1090$, and $\beta\_h = -0.0208$.

Figure 13: Crowding Externality and Optimal Crowding Tax: An Illustration



Notes: The diagram illustrates the market equilibrium, the social optimal, and the optimal crowding tax under the crowding externality.

The number of subway passengers in the market equilibrium is

$$N_{eq} = \text{avg crowding per minute} \times \text{CarArea} \times M = 4.16 \times 53.2 \times 6 = 1,328.$$

Given a subway share of 0.38, the total population is

$$POP = \frac{N_{eq}}{\text{share\_metro}} = \frac{1,328}{0.38} = 3495,$$

which is assumed to be evenly split between high- and low-income groups, so that $POP_l = POP_h = 1,748$. The congestion exposure is computed as

$$L = \frac{\text{avg TravelTime}_t}{\text{CarArea} \times K} = \frac{30.5}{53.2 \times 6} = 0.096,$$

and the operator price is set to $P_{\text{base}} = 5.03$. Finally, let the private demand function satisfy $PD(N) =$

P_{base} in equilibrium, we solve for $A = 1.43$.

7.3 Welfare under Alternative Policies

In this subsection, we evaluate three alternative policies designed to reduce subway crowding: (1) optimal crowding tax, (2) two-class configuration, and (3) quantity control at optimal ridership. We assess their welfare implications relative to the current EBD-induced market equilibrium, focusing on aggregate and distributional consumer surplus, operator's revenue, ridership and crowdedness, and changes in congestion on surface roads.

Current market equilibrium: For group g with two alternatives (metro vs. outside), EV1 implies

$$CS_g(N,P) \;=\; -\frac{POP_g}{\alpha_g}\,\ln\!\left(1+e^{A+\alpha_g P+\beta_g LN}\right), \qquad CS=\sum_g CS_g.$$

As discussed in Section 7.2, given the operator price P_{base}, estimated parameters α_g and β_g, calibrated POP_g, A and L, we can solve $PD(N) = P_{base}$ for N by bisection on N in $[0, M]$, and thus compute CS_g and the operator's revenue as $FareRev = P_{base}N$.

Pigouvian crowding tax: Given the operator price P_{base}, the constrained social optimum N^* can be obtained by solving $PD(N^*) - w(N^*) \;=\; P_{\text{base}}$, and the corresponding optimal crowding tax is $T^* \;=\; w(N^*)$. Riders pay $P_{\text{cons}} = P_{\text{base}} + T^*$. At (N^*, P_{cons}), private and social conditions coincide. The group-specific subway ridership and welfare can be evaluated as $N_g = POP_g s_g(A + \alpha_g P_{\text{cons}} + \beta_g L N^*)$ and $CS_g(N^*, P_{\text{cons}})$, respectively. And the tax revenue is given as $\text{TaxRev} = T^*N^*$.

Two-class configuration: When the train is split into two classes (3 business and 3 standard cars), utilities become

$$V_{gB} = A + \alpha_g P_B + \beta_g L_B N_B, \qquad V_{gS} = A + \alpha_g P_S + \beta_g L_S N_S,$$

with $L_B = L(6/3) = 2L$ and $L_S = 2L$. The three–alternative (B, S, Outside) shares are

$$(s_{gB}, s_{gS}, s_{gO}) \;=\; \left(\frac{e^{V_{gB}}}{1+e^{V_{gB}}+e^{V_{gS}}}, \frac{e^{V_{gS}}}{1+e^{V_{gB}}+e^{V_{gS}}}, \frac{1}{1+e^{V_{gB}}+e^{V_{gS}}}\right),$$

and counts $N_B = \sum_g POP_g s_{gB}$, $N_S = \sum_g POP_g s_{gS}$. For group g, EV1 implies

$$CS_g(N_B, N_S, P_B, P_S) \;=\; -\frac{POP_g}{\alpha_g}\,\ln\!\left(1+e^{V_{gB}}+e^{V_{gS}}\right), \qquad CS=\sum_g CS_g.$$

The operator's revenue is $R(P_B, P_S) \;=\; P_B N_B + P_S N_S$. Particularly, we consider two scenarios: (i) revenue-neutral with a fixed price ratio $P_B \;=\; \rho P_S$, and (ii) revenue-neutral with unconstrained welfare-maximized prices $P_B > P_S$.

Quantity control: If the metro is *available* with probability $q \in (0, 1]$ to everyone (ex–ante), availability enters as a multiplicative weight inside the log–sum, equivalently as an additive $\ln q$ to metro utility. Target the same total ridership N^* as the Pigouvian social optimal, the group-specific consumer surplus is written as:

$$CS_g^{\text{lot}}(N^*, P, q) \;=\; -\frac{POP_g}{\alpha_g}\,\ln\!\Big(1 + q\,e^{A+\alpha_g P+\beta_g L N^*}\Big).$$

External Road Congestion (MECC): Lastly, we consider the spillover impact of subway policies on surface road congestion. Let $\text{MECC}_{\text{veh-km}}$ be yuan per vehicle–kilometer and $\bar{d}$ the mean trip distance (km). With average car occupancy κ, the congestion cost per *passenger–km* in cars is $c_{\text{car}} = \text{MECC}_{\text{veh-km}}/\kappa$. If a rider shifts off the metro, we presumably assign the probabilities of taking other modes as

$$p_{\text{bus},\ell} = 0.5,\;\; p_{\text{car},\ell} = 0.25,\;\; p_{\text{other},\ell} = 0.25$$

$$p_{\text{bus},h} = 0.15,\;\; p_{\text{car},h} = 0.75,\;\; p_{\text{other},h} = 0.1$$

We assume that 1 bus per km produces 0.4 car per km of congestion following Gu et al. (2023). Assume that passengers shifting to other modes do not cause congestion to the surface road traffic. The expected congestion factor (in car–passenger–km equivalents) is

$$k_\ell \;=\; p_{\text{car},\ell} + 0.4\,p_{\text{bus},\ell}, \qquad k_h \;=\; p_{\text{car},h} + 0.4\,p_{\text{bus},h},$$

The change in MECC (vs. baseline) is

$$\Delta\text{MECC}_g \;=\; c_{\text{car}}\,k_g\,\bar{d}\,\big(N_g^{\text{base}} - N_g^{\text{cf}}\big), \qquad \Delta\text{MECC} = \sum_g \Delta\text{MECC}_g.$$

Welfare comparison: We calculate total welfare without MECC as $W = CS + \text{TaxRev}$, and with MECC as $W - \Delta\text{MECC}$. Under revenue neutrality, FareRev is fixed. Also note that the crowding externalities have been included in the log-sum consumer surpluses so that no need to deduct them repeatedly in the welfare calculation. We compare prices and ridership across scenarios in Table 14, and the corresponding results for revenue and welfare in Table 15.

The optimal crowding tax under the single-class configuration is 5.49 RMB, which is slightly higher than the baseline fare of 5.03 RMB. The higher tax-inclusive price reduces total subway ridership by about 28% relative to the competitive equilibrium. This reduction falls disproportionately on low-income passengers, whose ridership and welfare both decline substantially. In contrast, despite the higher price, more high-income passengers take the subway due to lower crowdedness and their utility also increases. In aggregate, in-system welfare - measured as consumer surplus plus tax revenue, with the fare assumed equal to marginal operating cost - increases by about 8%. However, much of these gains are offset by increased congestion externality on surface roads, as displaced subway riders shift to alternative modes.

The quantity control policy under the single-class configuration enforces the optimal level of ridership through a lottery mechanism. The lottery winning rate is 0.44, because more poeple will "apply" than the number of passengers in competitive equilibrium. Compared with the optimal crowding tax, quantity control is more favorable to low-income passengers, who enjoy more trips and higher welfare. Nevertheless, the overall consumer welfare is slightly lower and the aggregate system-wide welfare is far below that achieved with the crowding tax.

The third counterfactual introduces a two-class configuration and requires to generate the same revenue as in single-class competitive equilibrium. We focus on the scenario where prices are set such that $P_B > P_S$ and the in-system welfare is maximized. The equilibrium price for business class is 4.93 RMB, about 50% higher than in standard class (3.24 RMB), but both below the base fare of 5.03. This is because more passengers bring down fare level needed to generate the same revenue. High-income passengers self-select into pricier but less-crowded business class, while low-income passengers prefer the cheaper standard class. By inducing self-selection, it delivers a Pareto improvement for both income groups in terms of increased ridership and utility. In addition, by attracting more passengers to the subway, the external cost on surface road congestion reduces, further enhancing social welfare.

Table 14: Comparisons across Scenarios: Prices and Ridership

	P	tax	lottery prob.	N	N_l	N_h	k
1. Single-class config.							
1.1 Competitive equilibrium (status quo)	5.03	-	-	1,328	1,074	254	4.16
1.2 Optimal crowding tax	5.03	5.49	-	960	675	285	3.01
1.3 Quantity control (N under 1.2)	5.03	-	0.44	960	726	234	3.01
2. Two-class 3-3 config. (fare rev. in 1.1)							
2.1 $P_B = 5P_S$, All	4.00	-	-	1,671	1,368	303	5.23
Business	8.36	-	-	581	343	238	3.64
Standard	1.67	-	-	1,089	1,025	65	6.83
2.2 Welfare maximization	3.53	-	-	1,663	1,376	287	5.21
Business	4.93	-	-	763	594	169	4.78
Standard	3.24	-	-	900	781	118	5.64

Notes: Equilibrium ridership and crowding under alternative policies.

8 Conclusions

This paper provides the first revealed-preference estimation of passengers' monetary willingness to pay (WTP) to avoid subway crowding. On average, passengers are willing to pay 2.4 RMB to reduce

Table 15: Comparisons across Scenarios: Revenue and Welfare

(Unit, 1,000 RMB)	Consumer Surplus			Revenue		Δ Cong.	Welfare	
	All	CS_l	CS_h	Fare	Tax	Exter.	w/o ΔCE	w/ ΔCE
1. Single-class config.								
1.1 Competitive equilibrium (status quo)	12.1	9.6	2.5	4.8	0	0 (ref.)	12.1	12.1
1.2 Optimal crowding tax	7.8	4.9	2.9	4.8	5.3	0.8	13.0	12.2
1.3 Quantity control (N under 1.2)	7.7	5.4	2.3	4.8	0	0.9	7.7	6.8
2. Two-class 3-3 config. (fare rev. in 1.1)								
2.1 $P_B = 5P_S$	18.4	15.3	3.1	6.7	0	-0.9	18.4	19.3
2.2 Welfare maximization	18.5	15.6	2.9	6.7	0	-0.9	18.5	19.3

Notes: Aggregate and distributional welfare impacts under alternative policies.

crowding by one person per square meter for a 30-minute in-train trip with the travel distance of 20 km and fare of 5 RMB. THe marginal WTP increases with both the level of crowdedness and passenger income. We also document a sizable negative crowding externality, estimated at about 10 RMB per passenger.

Building on these estimates, we conduct welfare and counterfactual policy analyses. An optimal crowding tax of 5.5 RMB, while slightly improving in-system welfare, reduces equilibrium ridership by about 28% and disproportionately harms low-income passengers. By contrast, a two-class configuration with 2nd-degree price discrimination induces self-selection: high-income passengers opt into a less crowded, higher-priced subway car, while low-income passengers choose a cheaper standard one. This design results in higher utility for both income groups and increases overall welfare. A quantity control policy performs much worse, delivering lower consumer surplus and failing to generate tax revenues. Finally, we show that subway policies generate externalities on surface road congestion.

References

Arnott, R. (2013): "A bathtub model of downtown traffic congestion," *Journal of Urban Economics*, 76, 110–121.

Arnott, R., A. De Palma, and R. Lindsey (1993): "A structural model of peak-period congestion: A traffic bottleneck with elastic demand," *The American Economic Review*, 161–179.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

Chu, X. (1995): "Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach," *Journal of Urban Economics*, 37, 324–343.

Cook, C. and P. Z. Li (2025): "Value pricing or lexus lanes? the distributional effects of dynamic tolling," Tech. rep.

Coulombel, N. and G. Monchambert (2023): "Diseconomies of scale and subsidies in urban public transportation," *Journal of Public Economics*, 223, 104903.

Daganzo, C. F. (1994): "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation research part B: methodological*, 28, 269–287.

Davis, L. W. (2021): "Estimating the price elasticity of demand for subways: Evidence from Mexico," *Regional Science and Urban Economics*, 87, 103651.

Douglas, N. and G. Karpouzis (2006): "Estimating the passenger cost of train overcrowding," in *29th Australian Transport Research Forum*, 1–8.

Gu, Y., Q. Tang, Y. Wang, and B. Zou (2023): "Fare structure and the demand for public transit," *Work Pap.*

Hahn, R. W., R. D. Metcalfe, and E. Tam (2023): "Welfare Estimates of Shifting Peak Travel," Tech. rep., National Bureau of Economic Research.

Hall, J. D. (2024): "Inframarginal travelers and transportation policy," *International Economic Review*, 65, 1519–1550.

Haywood, L. and M. Koning (2015): "The distribution of crowding costs in public transport: New evidence from Paris," *Transportation Research Part A: Policy and Practice*, 77, 182–201.

Henderson, J. V. (1974): "Road congestion: a reconsideration of pricing theory," *Journal of Urban Economics*, 1, 346–365.

Hörcher, D., D. J. Graham, and R. J. Anderson (2017): "Crowding cost estimation with large scale smart card and vehicle location data," *Transportation Research Part B: Methodological*, 95, 105–125.

Kreindler, G. (2024): "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *Econometrica*, 92, 1233–1268.

Li, Z. and D. A. Hensher (2011): "Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal," *Transport Policy*, 18, 880–887.

Lighthill, M. J. and G. B. Whitham (1955): "On kinematic waves II. A theory of traffic flow on long crowded roads," *Proceedings of the royal society of london. series a. mathematical and physical sciences*, 229, 317–345.

Lu, H., T. Fowkes, and M. Wardman (2008): "Amending the incentive for strategic bias in stated preference studies: case study in users' valuation of rolling stock," *Transportation research record*, 2049, 128–135.

Parry, I. W. H. and K. A. Small (2009): "Should urban transit subsidies be reduced?" *American Economic Review*, 99, 700–724.

Richards, P. I. (1956): "Shock waves on the highway," *Operations research*, 4, 42–51.

Singh, J., G. H. de Almeida Correia, B. van Wee, and N. Barbour (2023): "Change in departure time for a train trip to avoid crowding during the COVID-19 pandemic: A latent class study in the Netherlands," *Transportation Research Part A: Policy and Practice*, 170, 103628.

Vickrey, W. S. (1969): "Congestion theory and transport investment," *The American economic review*, 59, 251–260.

Wardman, M. and G. Whelan (2011): "Twenty years of rail crowding valuation studies: evidence and lessons from British experience," *Transport reviews*, 31, 379–398.

Yap, M., O. Cats, and B. Van Arem (2020): "Crowding valuation in urban tram and bus transportation based on smart card data," *Transportmetrica A: Transport Science*, 16, 23–42.

Crowding
Online Appendix (Not for Publication)

Table of Contents

A Robustness Checks **2**

A.1 Varying Definitions of Accidental Companions 2

A.2 Adding Line Fixed Effects 3

A.3 Including Origin Stations 4

A.4 ACP as IV 7

A.5 Over-identification in Standard Logit 9

A.6 Crowding and Delay 11

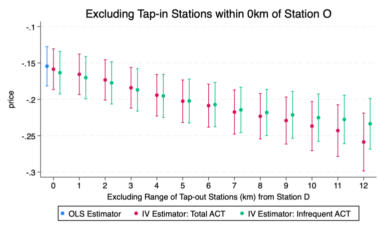
B Additional Figures and Tables **14**

A Robustness Checks

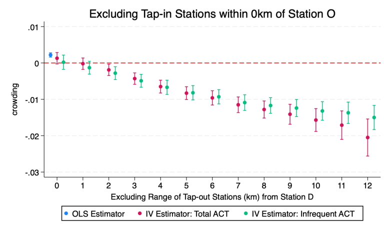
A.1 Varying Definitions of Accidental Companions

Figure A.1: 2SLS Estimation with Different Construction of IVs

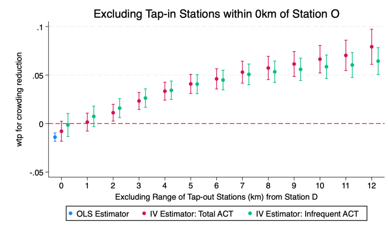
Panel A: Accidental Companions with Tap-in Stations other than O



(a) α

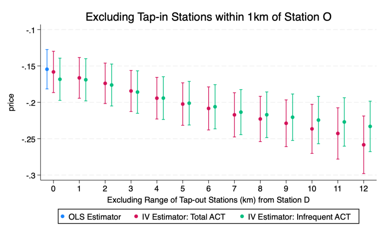


(b) β

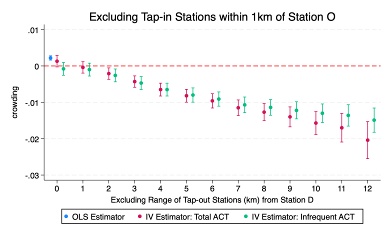


(c) WTP

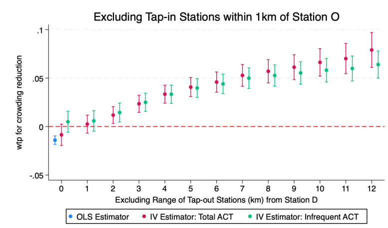
Panel B: Accidental Companions with Tap-in Stations more than 1 km away from O



(d) α

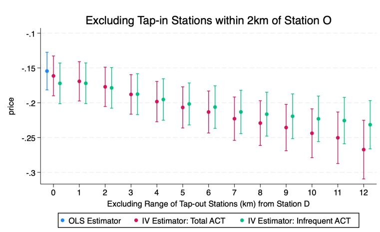


(e) β

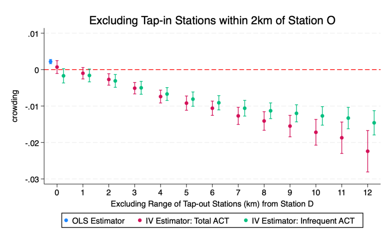


(f) WTP

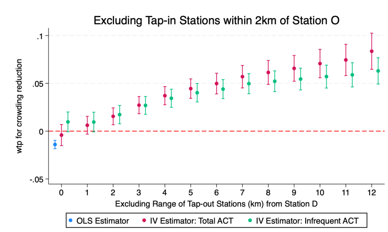
Panel C: Accidental Companions with Tap-in Stations more than 2 km away from O



(g) α



(h) β



(i) WTP

Notes: Coefficients associated with crowding, rescheduling, and the WTPs from OLS and GMM estimates where ACTs are constructed in various ways.

A.2 Adding Line Fixed Effects

Table A.1: Adding Destination Line-by-Time Bin Fixed Effects

	All Riders			Frequent Riders		
	Log share	Crowding	Log share	Log share	Crowding	Log share
	(1)	(2)	(3)	(4)	(5)	(6)
Price	-0.1804	-4.1729	-0.2428	-0.1843	-4.2050	-0.2490
	(0.0140)	(0.5154)	(0.0167)	(0.0147)	(0.5316)	(0.0175)
Rescheduling	-0.0243	-0.3319	-0.0285	-0.0257	-0.3489	-0.0302
	(0.0005)	(0.0192)	(0.0007)	(0.0006)	(0.0199)	(0.0008)
Crowding	0.0020		-0.0136	0.0021		-0.0139
	(0.0003)		(0.0013)	(0.0003)		(0.0013)
ACT		9.1380			9.3924	
		(0.3493)			(0.3631)	
OD-Day FE	X	X	X	X	X	X
Time-Destin Line FE	X	X	X	X	X	X
Model	OLS	First Stage	2SLS	OLS	First Stage	2SLS
Observations	200,624	200,624	200,624	187,928	187,928	187,928
R^2	0.7228			0.7179		
First Stage F-Stat		684.3	684.3		669	669
Implied Willingness-to-Pay (WTP)						
Mean Rescheduling	15.02		15.02	14.90		14.90
Mean Crowding	127.4		127.4	127.4		127.4
MWTP for Resched. Avoidance	0.1350		0.1170	0.1390		0.1210
	(0.0109)		(0.0083)	(0.0116)		(0.0087)
MWTP for Crowding Reduction	-0.0110		0.0562	-0.0113		0.0559
	(0.0018)		(0.0057)	(0.0019)		(0.0057)

Notes: Line of origin-by-time bin fixed effects cannot be included as they absorb price variation created by the EBD.

A.3 Including Origin Stations

Table A.2: OLS Regressions: Including the Origin Stations

	Sample: overall		Sample: frequent riders	
	log_share	log_share	log_share	log_share
	(1)	(2)	(3)	(4)
price	-0.0989***	-0.0979***	-0.0993***	-0.0981***
	(0.0127)	(0.0127)	(0.0134)	(0.0134)
rescheduling	-0.0248***	-0.0249***	-0.0263***	-0.0265***
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
crowding	0.0028***	0.0029***	0.0029***	0.0030***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Observations	236,952	236,952	221,568	221,568
R^2	0.7359	0.7366	0.7334	0.7341
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X
Mean rescheduling	15.22	15.22	15.07	15.07
Mean crowding	128.6	128.6	128.3	128.3
MWTP for resheduling avoidence	0.251***	0.255***	0.265***	0.27***
	(0.0335)	(0.0343)	(0.0370)	(0.0381)
MWTP for crowding reduction	-0.0282***	-0.0295***	-0.0294***	-0.031***
	(0.0045)	(0.0047)	(0.0048)	(0.0051)

Notes: Robustness checks with the line-origin stations included (four more origin stations in the sample). OLS regressions.

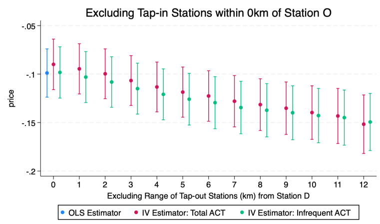
Table A.3: GMM2S: Including the Origin Stations: 2nd Stage

	Sample: overall		Sample: frequent riders	
	log_share	log_share	log_share	log_share
	(1)	(2)	(3)	(4)
price	-0.1372***	-0.1391***	-0.1405***	-0.1423***
	(0.0140)	(0.0141)	(0.0148)	(0.0150)
rescheduling	-0.0279***	-0.0278***	-0.0296***	-0.0296***
	(0.0006)	(0.0006)	(0.0007)	(0.0007)
crowding	-0.0080***	-0.0084***	-0.0083***	-0.0086***
	(0.0010)	(0.0010)	(0.0010)	(0.0011)
First Stage F-Test	886.7	782	868	765.6
Observations	236,952	236,952	221,568	221,568
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X
Mean rescheduling	15.22	15.22	15.07	15.07
Mean crowding	128.6	128.6	128.3	128.3
MWTP for resheduling avoidence	0.203***	0.200***	0.211***	0.208***
	(0.0213)	(0.0209)	(0.0227)	(0.0223)
MWTP for crowding reduction	0.0585***	0.0601***	0.0592***	0.0604***
	(0.0082)	(0.0084)	(0.0083)	(0.0085)

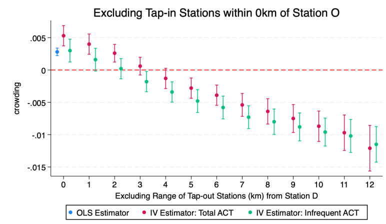
Notes: Robustness checks with the line-origin stations included (four more origin stations in the sample). Second-stage regressions.

Figure A.2: 2SLS Estimations with Varying Definitions of ACT: Including the Origin Stations

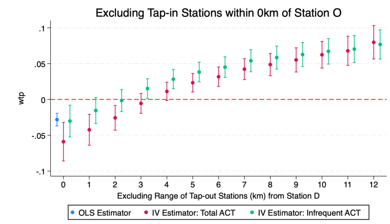
Panel A: Accidental Companions with Tap-in Stations other than O



(a) α

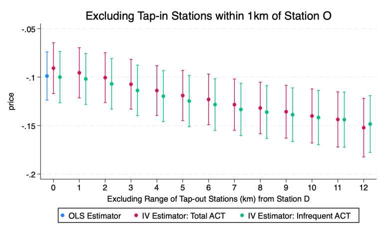


(b) β

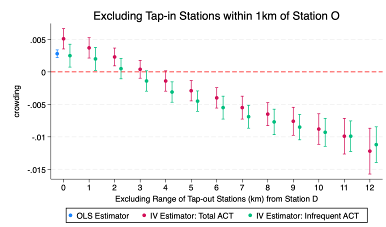


(c) WTP

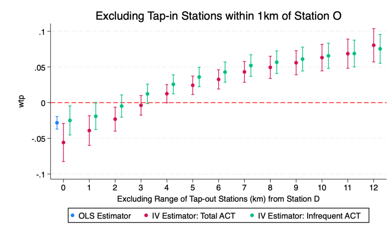
Panel B: Accidental Companions with Tap-in Stations more than 1 km away from O



(d) α

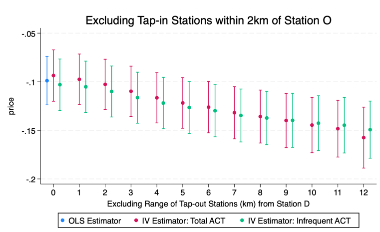


(e) β

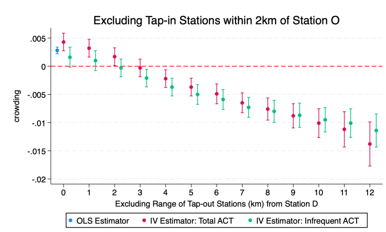


(f) WTP

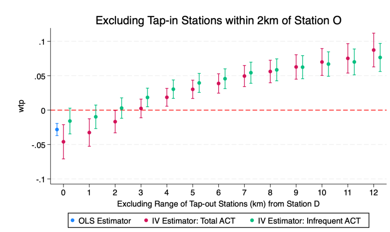
Panel C: Accidental Companions with Tap-in Stations more than 2 km away from O



(g) α



(h) β



(i) WTP

Notes: Robustness checks with the line-origin stations included (four more origin stations in the sample). Coefficients associated with crowding, rescheduling, and the WTPs from OLS and GMM estimates where ACTs are constructed in various ways.

A.4 ACP as IV

Table A.4: First Stage Single IV GMM Estimates (IV: ACP)

	Sample: overall		Sample: frequent riders	
	crowding	crowding	crowding	crowding
	(1)	(2)	(3)	(4)
price	-4.6900	-4.8148	-4.8486	-4.9843
	(0.5273)	(0.5257)	(0.5460)	(0.5440)
rescheduling	-0.3109	-0.2905	-0.3275	-0.3053
	(0.0202)	(0.0203)	(0.0209)	(0.0211)
ACP	158.1939	157.4474	162.5137	161.4234
	(7.2581)	(7.5171)	(7.5109)	(7.7684)
First Stage F-Test	475	438.7	468.2	431.8
Observations	200,624	200,624	187,928	187,928
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X

Notes: Using accidental companion passengers (ACPs) instead of accidental companion time (ACTs) as the instrumental variable for crowding. First-stage estimations.

Table A.5: Second Stage Single IV GMM Estimates (IV: ACP)

	Sample: Overall		Sample: Frequent Riders	
	(1)	(2)	(3)	(4)
Price	-0.2243	-0.2280	-0.2294	-0.2332
	(0.0170)	(0.0173)	(0.0180)	(0.0183)
Rescheduling	-0.0287	-0.0286	-0.0305	-0.0303
	(0.0007)	(0.0007)	(0.0008)	(0.0008)
Crowding	-0.0130	-0.0135	-0.0133	-0.0138
	(0.0013)	(0.0014)	(0.0014)	(0.0015)
First Stage F-Test	475.0	438.7	468.2	431.8
Observations	200,624	200,624	187,928	187,928
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X
Implied Willingness-to-Pay (WTP)				
Mean Rescheduling	15.02	15.02	14.90	14.90
Mean Crowding	127.4	127.4	127.4	127.4
MWTP for Resched. Avoidance	0.1280	0.1250	0.1330	0.1300
	(0.0100)	(0.0098)	(0.0106)	(0.0104)
MWTP for Crowding Reduction	0.0580	0.0591	0.0581	0.0591
	(0.0060)	(0.0062)	(0.0061)	(0.0062)

Notes: Using accidental companion passengers (ACPs) instead of accidental companion time (ACTs) as the instrumental variable for crowding. Second-stage estimations.

A.5 Over-identification in Standard Logit

Table A.6: GMM2S: Multiple IV & 1st Stage

	Sample: overall		Sample: frequent riders	
	crowding	crowding	crowding	crowding
	(1)	(2)	(3)	(4)
price	42.5050***	41.7337***	42.5962***	41.8534***
	(2.3341)	(2.3422)	(2.4294)	(2.4350)
rescheduling	-0.3693***	-0.3478***	-0.3834***	-0.3600***
	(0.0175)	(0.0177)	(0.0182)	(0.0184)
ACT	9.4848***	9.2742***	9.7859***	9.5560***
	(0.3666)	(0.3843)	(0.3826)	(0.4006)
treat × before 7 am	73.0341***	71.7530***	72.8614***	71.6321***
	(3.7644)	(3.7773)	(3.9159)	(3.9246)
treat × time trend	659.1104***	663.6061***	665.4799***	670.0627***
	(35.6669)	(35.6264)	(36.8151)	(36.7568)
treat × time trend^2	-22.0289***	-22.1823***	-22.2570***	-22.4135***
	(1.1955)	(1.1940)	(1.2340)	(1.2319)
treat × time trend^3	0.2433***	0.2450***	0.2459***	0.2477***
	(0.0133)	(0.0133)	(0.0138)	(0.0137)
First Stage F-Test	328.9	304.7	319.8	296.7
Observations	200,624	200,624	187,928	187,928
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X

Notes: Robustness checks with additional instrumental variables to over-identify the standard logit model. The same set of instrumental variables is used to identify the random coefficient model. First-stage estimations.

Table A.7: GMM2S: Multiple IV & 2nd Stage

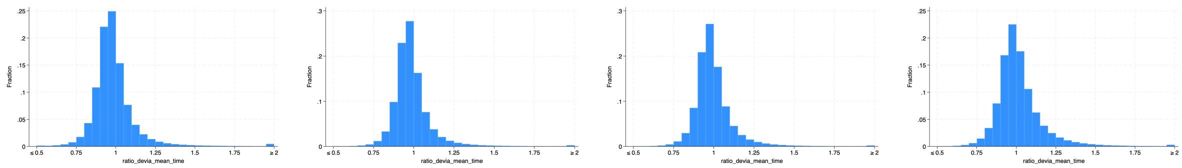
	Sample: overall		Sample: frequent riders	
	log_share	log_share	log_share	log_share
	(1)	(2)	(3)	(4)
price	-0.1583***	-0.1586***	-0.1640***	-0.1642***
	(0.0103)	(0.0103)	(0.0109)	(0.0109)
rescheduling	-0.0273***	-0.0273***	-0.0290***	-0.0289***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
crowding	-0.0065***	-0.0064***	-0.0066***	-0.0065***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)
First Stage F-Test	328.9	304.7	319.8	296.7
Observations	200,624	200,624	187,928	187,928
OD-Day FE	X	X	X	X
Time FE	X		X	
Time-Day FE		X		X
Mean rescheduling	15.02	15.02	14.90	14.90
Mean crowding	127.4	127.4	127.4	127.4
MWTP for resheduling avoidence	0.173***	0.172***	0.177***	0.176***
	(0.0118)	(0.0117)	(0.0122)	(0.0122)
MWTP for crowding reduction	0.041***	0.0403***	0.0404***	0.0395***
	(0.0055)	(0.0055)	(0.0055)	(0.0056)

Notes: Robustness checks with additional instrumental variables to over-identify the standard logit model. The same set of instrumental variables is used to identify the random coefficient model. Second-stage estimations.

A.6 Crowding and Delay

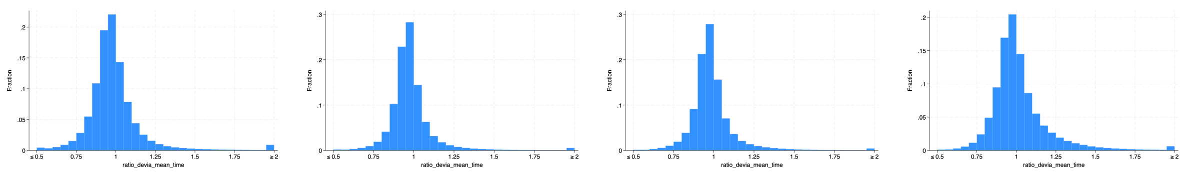
Figure A.3: Histograms of Ratio Deviation from the Mean Travel Time: by Quartiles of Crowding Density

Panel A: All Trips



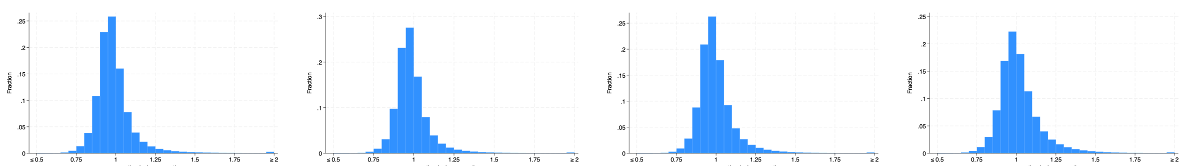
(a) Q1 (b) Q2 (c) Q3 (d) Q4

Panel B: Trips without Transfer



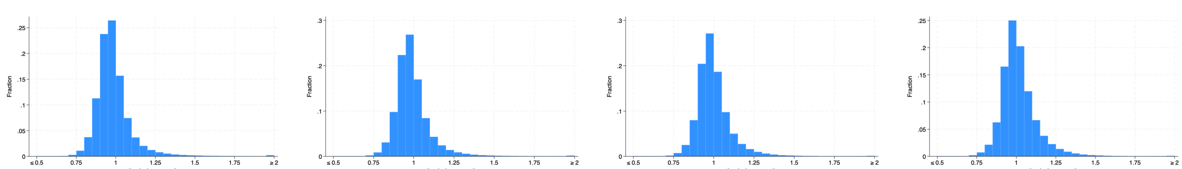
(e) Q1 (f) Q2 (g) Q3 (h) Q4

Panel C: Trips with One Transfer



(i) Q1 (j) Q2 (k) Q3 (l) Q4

Panel D: Trips with Two Transfers

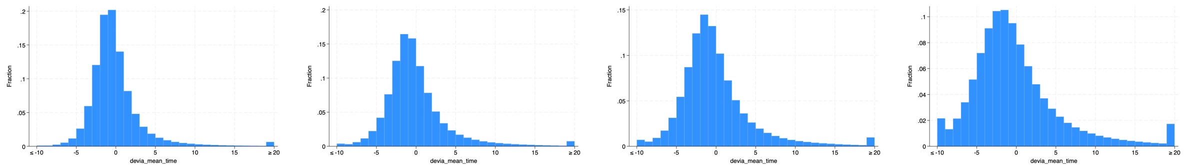


(m) Q1 (n) Q2 (o) Q3 (p) Q4

Notes: The graphs show the residualized distribution (from OD pair fixed effects) of travel time deviation. Panel A includes trips that require no transfer; Panel B includes trips that require one transfer; Panel C includes trips that require two transfers. In each panel, average crowding in the lowest quartile is on the left, that in the highest quartile is on the right.

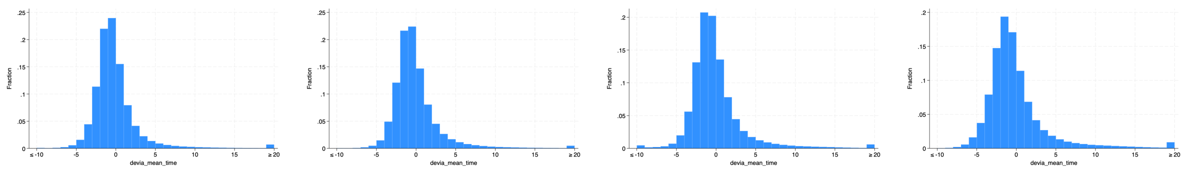
Figure A.4: Histograms of Deviation from the Mean Travel Time: by Quartiles of Travel Distance

Panel A: All Trips



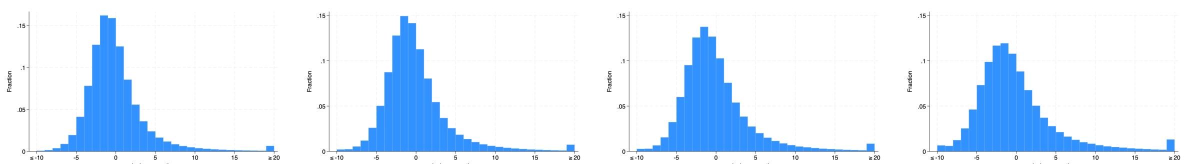
(a) Q1 (b) Q2 (c) Q3 (d) Q4

Panel B: Trips without Transfer



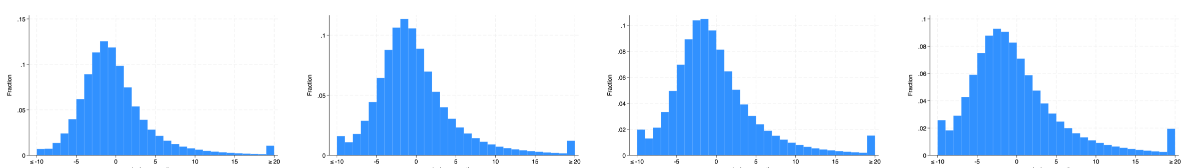
(e) Q1 (f) Q2 (g) Q3 (h) Q4

Panel C: Trips with One Transfer



(i) Q1 (j) Q2 (k) Q3 (l) Q4

Panel D: Trips with Two Transfers

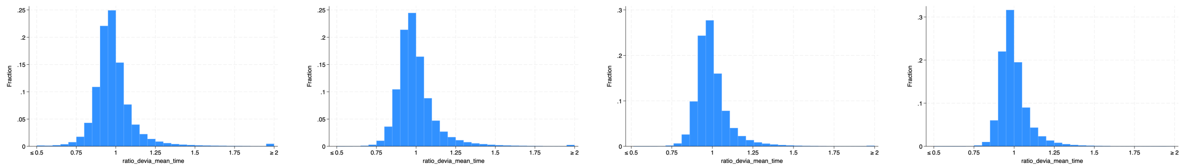


(m) Q1 (n) Q2 (o) Q3 (p) Q4

Notes: The graphs show the residualized distribution (from OD pair fixed effects) of travel time deviation. Panel A includes trips that require no transfer; Panel B includes trips that require one transfer; Panel C includes trips that require two transfers. In each panel, average travel time in the lowest quartile is on the left, that in the highest quartile is on the right.

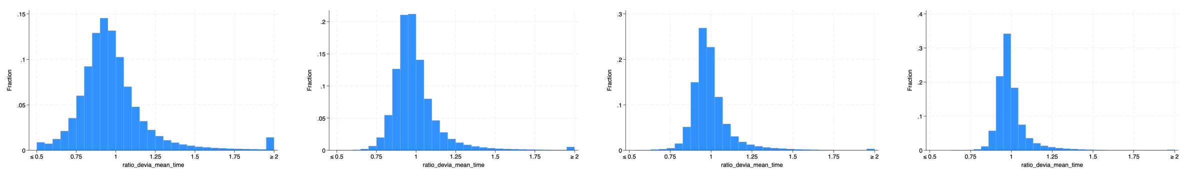
Figure A.5: Histograms of Ratio Deviation from the Mean Travel Time: by Quartiles of Travel Distance

Panel A: All Trips



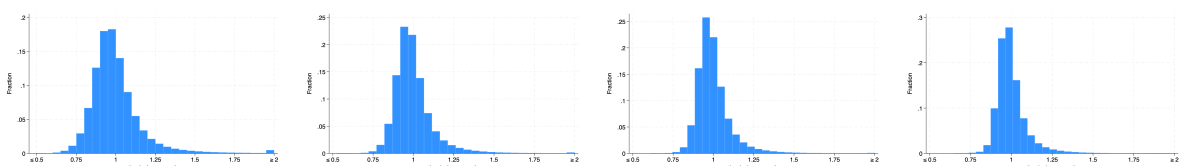
(a) Q1 (b) Q2 (c) Q3 (d) Q4

Panel B: Trips without Transfer



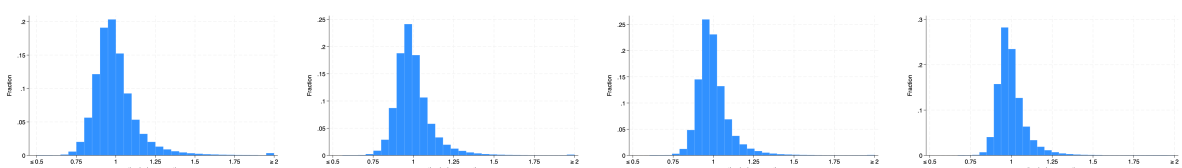
(e) Q1 (f) Q2 (g) Q3 (h) Q4

Panel C: Trips with One Transfer



(i) Q1 (j) Q2 (k) Q3 (l) Q4

Panel D: Trips with Two Transfers

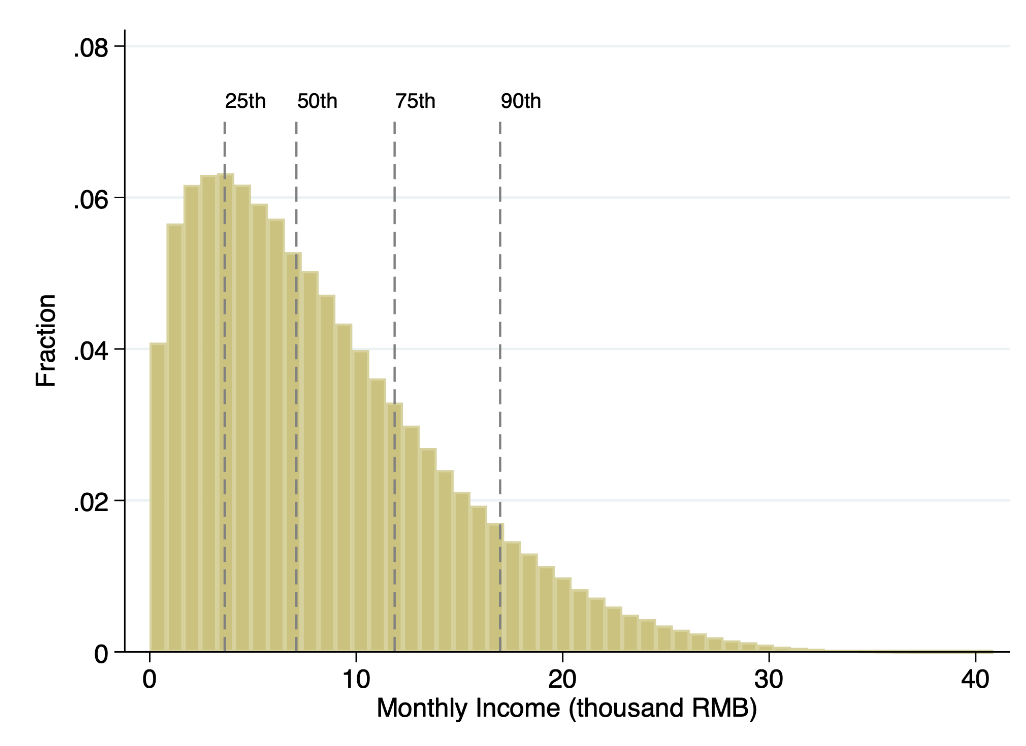


(m) Q1 (n) Q2 (o) Q3 (p) Q4

Notes: The graphs show the residualized distribution (from OD pair fixed effects) of travel time deviation, as a ratio of the mean travel time. Panel A includes trips that require no transfer; Panel B includes trips that require one transfer; Panel C includes trips that require two transfers. In each panel, average travel time in the lowest quartile is on the left, that in the highest quartile is on the right.

B Additional Figures and Tables

Figure B.1: Histogram of Simulated Monthly Incomes

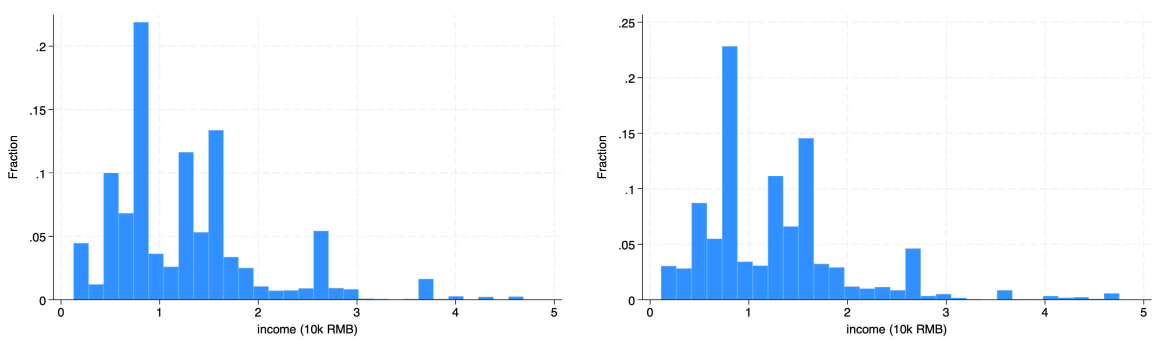


Note: Unit 10 thousand RMB.

Table B.1: Income Distribution by Subway Trip Types

	Income (10k RMB)	
	Commuting Trips	Other Trips
	(1)	(2)
Mean	1.31	1.30
SD	0.94	0.91
P1	0.21	0.23
P10	0.52	0.52
P25	0.75	0.76
P50	1.18	1.22
P75	1.56	1.56
P90	2.60	2.39
P99	4.69	4.69
N	4496	12835

Figure B.2: Income Histograms by Subway Trip Types



(a) Commuting (b) Other

Figure B.3: 8-km Radius in the Beijing Subway Network