Crowding

Yizhen Gu Qu Tang Kai Wu Ben Zou*

April 2026

Abstract

Crowding is a common disamenity, yet little is known about its monetary cost. We estimate the willingness to pay (WTP) to avoid crowding in public transit with a revealed-preference approach. Beijing Subway passengers choose departure times, trading off fare, in-train crowding, and schedule deviation, with price variation from an early-bird discount. Leveraging the subway's network structure, we impute real-time in-train crowding, infer each passenger's optimal arrival time, and construct a novel instrumental variable for crowding. The marginal WTP to reduce crowding by one passenger per square meter is about 0.05 RMB per minute, implying a crowding externality exceeding the fare. High-income passengers have higher crowding WTP but are less price-sensitive. An optimal crowding tax improves welfare but is regressive; a two-class configuration elicits self-selection and benefits both high and low income commuters.

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.

*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.

There are at least three empirical challenges to credibly estimating the disutility from crowding. First, crowding is endogenous. It emerges when many people respond to the same, often unobserved, demand shock and converge on the same location at the same time. Second, crowding is correlated with other undesired features that are unobservable to the researcher. 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 at higher prices, such as business classes on airplanes and skyboxes in stadiums, and those options typically come with other perks, such as better services or a symbol of status. In some contexts, crowding may even 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. What people *state*, however, may differ from how they actually *behave*.

This paper estimates the marginal willingness to pay (MWTP) to avoid crowding in public transit, focusing on the Beijing Subway. Crowding is an important supply-side challenge in many transit systems, and quantifying passengers' MWTP to avoid it provides a key input for service design and pricing. We leverage an off-peak price discount introduced in December 2015 and exploit the resulting variation in ticket prices across departure times. The discount, known as the Early-Bird Discount (EBD), gives a 30 percent reduction to trips that originate from 16 stations on two suburban lines and tap in 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 depart before the EBD cutoff to receive the lower fare. Second, the passenger wishes to avoid an overly crowded train. Third, the passenger has an optimal arrival time, and any deviation from that optimal time is costly.

Rather than asking passengers to report subjective evaluations, we build a discrete-choice model of travel demand that backs out preferences for price, crowding, and schedule deviation from observed choices. Crucially, the extent to which passengers pay a higher price for a less crowded train reveals their WTP to avoid crowding in monetary terms. A distinctive advantage of the Beijing Subway is that crowding there can be treated as a nearly pure discomfort. Subway cars are otherwise identical; trains run on dedicated tracks and mostly on time; and we show empirically that crowding does not meaningfully slow down trains or induce detour routing. These features rule out the most prominent confounders that complicate crowding valuation in buses, highways, and other shared spaces.

The primary data source is the universe of trip-level records captured by smartcards in the Beijing Subway. The sample includes morning peak-hour trips on 21 workdays in 2016 that originate from the 16 EBD stations and from 16 comparison stations on two other suburban lines with similar ridership patterns. The trip records contain only the origin and destination stations and the tap-in and tap-out timestamps. To recover experienced crowding, we need to assign each passenger to a

specific train on a specific segment at a specific moment. We split each observed travel time into an in-train component and in-station components (walking in to the platform, waiting for the train, transferring, and walking out), and devise an iterative algorithm that jointly estimates each passenger's fastest route and the average access, transfer, and exit time at each station, with in-train time taken from the published timetable. Once the algorithm converges, we know every passenger's real-time position on the network. Dividing the number of passengers sharing the same train segment in the same five-minute bin by the available train floor area yields an instantaneous crowding density measured in persons per square meter, and integrating this density over the duration of the trip yields each passenger's total in-train crowding experience in person-minutes per square meter. Trains in Beijing's morning rush hour are crowded: among trips originating from the 32 selected stations in our main analysis, the average density is about 4 persons per square meter and the ninetieth percentile reaches roughly 5.7 persons per square meter. We proxy passenger expectations of crowding by fitted values from a regression of realized crowding on OD-day-of-week-time-bin fixed effects, which capture the systematic component of crowding that experienced commuters can anticipate.

Despite the discomfort from crowding, passengers converge on the same trains because they respond to common demand shocks, such as conventional work hours that require arrival by 9 AM. To break this endogeneity, we construct an instrumental variable based on the notion of *accidental companion passengers* (ACPs). For a focal passenger 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', yet whose trip overlaps with the focal passenger on a shared train segment because of the configuration of the network. We further restrict ACPs to riders who take the subway infrequently and whose destinations lie at least eight kilometers from D, ensuring that their presence on the focal train reflects demand shocks unrelated to the focal market. The instrument for crowding is the weighted sum of ACPs, where each companion is weighted by the duration of shared travel. We call this measure *accidental companion time* (ACT). ACT passes conventional weak-instrument tests with first-stage Wald F-statistics well above 280. The estimated crowding coefficient is wrong-signed under OLS, underscoring the importance of the ACT instrument. We show that the IV estimate of the MWTP to reduce crowding stabilizes as the spatial exclusion zone around D widens, consistent with the instrument eliminating destination-side demand correlation. The estimates are also stable across origin-side subsamples, reinforcing the argument that the relevant unobserved demand shocks are destination-driven.

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 arrive at the same destination around the same time on the same day but live elsewhere and originate from stations unaffected by the EBD. Because coworkers face no fare incentive to shift their schedules, their arrival distribution reflects unconstrained preferences. We perform quantile matching within each destination-date cell, mapping each EBD rider to the percentile rank of the control distribution

that she occupies among EBD arrivals. The method is non-parametric and exploits only contemporaneous data, avoiding concerns about pre-period drifts in commuting patterns, and assigns coworkers through two *distributions* rather than by guessing the workplace of each individual rider.

We find that passengers have a substantial MWTP to avoid crowding. The average passenger is willing to pay about 5 to 6 cents per minute to reduce crowding by one additional person per square meter. For a representative trip of 32 minutes (the ridership-weighted mean in-train time) at the mean crowding level, this amounts to roughly 1.8 RMB, or about 35 percent of the average fare of 5 RMB. Aggregated over all riders on a typical trip, the implied crowding externality exceeds the fare itself, indicating that unpriced crowding is a first-order distortion in subway pricing. The MWTP rises with the level of crowding, so welfare costs are concentrated on the most congested trips. A random-coefficient specification reveals that the MWTP for crowding reduction is steeply increasing in income, because higher-income riders both dislike crowding more and are less sensitive to fares. At the sample median income, the MWTP lies close to the homogeneous estimate; at the ninetieth percentile, it is more than twenty times larger. The MWTP to avoid rescheduling, by contrast, is approximately 0.10 RMB per minute and essentially *flat* across the income distribution. High-income and low-income passengers place similar value on arriving on time, even as they differ markedly in how much they will pay to avoid a crowded car. This divergence between the two income gradients turns out to be central to the welfare analysis.

Building on these estimates, we investigate the welfare consequences of three alternative policies designed to reduce crowding: a Pigouvian crowding tax set at the marginal externality, a capacity-rationing queue that rations access through waiting time, and a two-class configuration in which the existing six-car train is partitioned into a higher-priced Business section and a discounted Standard section under revenue neutrality. The crowding tax internalizes the externality directly, lowering ridership to the socially optimal level and raising total welfare relative to the unpriced equilibrium. Almost all of the welfare gain, however, is accounted for by the tax revenue itself, and the incidence is sharply regressive. Low-income riders bear the bulk of the welfare loss, while high-income riders gain slightly even before the revenue is rebated, because they value the reduction in crowding more than they dislike the higher fare. The queue targets the same ridership reduction but delivers lower welfare because it generates no transferable revenue. The queuing deadweight loss is almost exactly equal to the revenue the Pigouvian tax would collect, a classical equivalence between rationing by time and rationing by price in which the wait dissipates as deadweight loss what the tax would have captured as a transfer.

The two-class configuration, by contrast, delivers a Pareto improvement. Under revenue-neutral pricing at our preferred nesting parameter, the optimal Business fare is roughly 10.3 RMB and the optimal Standard fare is 2.4 RMB. Higher-income passengers self-select into the less crowded Business cars, the discounted Standard fare draws new riders from the outside option, and total ridership rises by about 18 percent. Variety-adjusted consumer surplus increases by roughly 8 percent, and

accounting for reduced road congestion from diverted car and bus trips pushes total welfare about 15 percent above baseline. Both income groups are strictly better off in net welfare terms, making two-class pricing the only instrument under which no group bears a welfare loss. The intuition is that two-class pricing exploits the steep income gradient in crowding WTP that is absent in the rescheduling dimension. Charging more for less crowded service aligns the fare gradient with the preference gradient, while the flat rescheduling WTP implies that alternative ways of differentiating the product, such as making Standard riders wait longer, would have much weaker distributional appeal.

Contributions to the literature. This paper contributes to several literatures. First, a large body of theoretical work suggests that users' responses to prices, road congestion, and crowding within the transit system are central considerations in the design of transit provision (e.g., Parry and Small, 2009; Coulombel and Monchambert, 2023). A growing empirical literature studies demand elasticity in public transit (e.g., Davis, 2021; Gu et al., 2023; Hahn et al., 2023) and road congestion, but the study of within-system crowding remains limited. Our paper is among the first to credibly estimate the monetary disutility from crowding. We emphasize that the subway setting differs conceptually from the classical literature on road congestion (e.g., Vickrey, 1969). Road congestion primarily generates externalities through lost time, whereas crowding in public transit generates disutility through discomfort.

Most existing studies of crowding 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 suffer from hypothetical bias and lack external validity. Some recent studies estimate passenger preferences from observed route choices using detailed trip-level records. Hörcher et al. (2017) studies the Hong Kong Subway and Yap et al. (2020) uses Dutch data. Both papers model discrete route choice where some routes cost less time but are more crowded and translate the estimated MWTP from time units into money using a rule-of-thumb value of time. We instead build an economic model that incorporates price, crowding, and schedule deviation jointly and provide the first revealed-preference estimate of WTP to avoid crowding directly in monetary terms, exploiting exogenous fare variation generated by the EBD. We further address the endogeneity of crowding with a novel instrument and estimate preference heterogeneity by income through a random-coefficient specification.

Second, we contribute to the study of optimal pricing under congestion externalities by providing the first empirical estimate of an optimal crowding tax and by comparing it against alternative instruments that discipline access through quantity rather than price. The welfare comparison connects to the long-standing discussion of price versus quantity controls (Weitzman, 1974; Li, 2018). Under heterogeneous disutility from crowding, we show that second-degree price discrimination through a two-class configuration elicits self-selection and produces a Pareto improvement, a result that complements Cook and Li (2025) on high-occupancy toll lanes. Our ACT instrument is struc-

turally analogous to the capacity-cost instrument of Hortaçsu et al. (2024), who use demand from other itineraries sharing the same flight leg to instrument for endogenous airline fares, and shares the logic that identifying variation in shared-infrastructure markets can come from travelers who use the same physical capacity but serve distinct markets.

Third, we develop new measurement tools that exploit the granular trip-level data and the network structure of the subway system. Although trip-level transit records are increasingly used in empirical work (e.g., Gu et al., 2023), they typically contain only entry and exit times and stations and do not allow researchers to trace passengers through the network. We develop an iterative algorithm that imputes each passenger's optimal route, real-time location, and in-train crowding. We also propose a novel way of measuring the optimal arrival time, a crucial but typically unobserved input in a large class of structural travel models with endogenous congestion, including 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). Our approach constructs coworkers among riders unaffected by the EBD and recovers their arrival preferences nonparametrically through quantile matching. We contribute to recent studies that estimate scheduling flexibility among commuters (e.g., Gu et al., 2023; Kreindler, 2024; Hall, 2024). Together, these tools allow us to measure crowding and rescheduling disutility with greater precision and may prove useful in other settings with similarly granular data.

The rest of the paper is organized as follows. Section 2 introduces the demand model and discusses the empirical challenges of measurement and endogeneity. Section 3 describes the institutional setting, the data, and the construction of crowding, optimal arrival times, the ACT instrument, income, and market size. Section 4 documents stylized facts that motivate the modeling choices. Section 5 presents the estimation results for both the homogeneous logit and the random-coefficient specifications. Section 6 conducts the welfare and counterfactual analysis. Section 7 concludes.

2 Demand Model

This section presents a model of travel demand in which subway passengers choose their departure time, trading off the fare, expected in-train crowding, and the cost of deviating from an optimal schedule. We first describe the model and then discuss the empirical challenges to estimation.

2.1 A Structural Model of Subway Demand with Crowding

We define a *market* as an origin-destination (OD) station pair j on a workday d between 6:30–8:30 AM measured by the departure time (tap-in time at the station of origin). Each market offers a set of *products* indexed by 15-minute tap-in time bins $t \in \mathcal{T}$. The outside option ($t = 0$) represents

traveling between the same origin and destination during the morning rush hour by other modes.

Passenger i's utility from departing at time t in market (j, d) is

$$U_{ijdt} = V_{ijdt} + \varepsilon_{ijdt},, \tag{1}$$

where the error term, ε_{ijdt}, is i.i.d. following a Type-I extreme value distribution, and the deterministic component is

$$V_{ijdt} = \alpha_i P_{jdt} + \beta_i \, \mathbb{E}_d[\text{Crowd}_{jdt}] + \rho_i \left| t + \mathbb{E}_d[\text{TT}_{jdt}] - t^{OA}_{ijd} \right| + \kappa_{jd} + \phi_t + \xi_{jdt}. \tag{2}$$

The fare P_{jdt} varies across OD pairs as a step function of distance and varies across time bins within an OD pair at EBD stations, where passengers receive a 30% discount for tapping in before 7 AM. The term $\mathbb{E}_d[\text{Crowd}_{jdt}]$ is the expected total in-train crowding experienced by a passenger during the trip, measured in passengers · minutes per square meter; it aggregates the instantaneous crowding density (passengers per square meter) over the duration of the passenger's total in-train time. The rescheduling cost $\left| t + \mathbb{E}_d[\text{TT}_{jdt}] - t^{OA}_{ijd} \right|$ captures the absolute deviation between the passenger's actual arrival time—the sum of the departure time t and expected travel time $\mathbb{E}_d[\text{TT}_{jdt}]$—and the optimal arrival time t^{OA}_{ijd}. We assume the penalty from deviating from the optimal arrival time is symmetric and linear in time, which simplifies the estimation while keeping the focus on the price-crowding tradeoff.[1] The remaining terms are OD-day fixed effects κ_{jd}, which absorb the common value of completing trip j on day d; time-bin fixed effects ϕ_t, which capture common intraday demand patterns across all markets; and a structural demand shock ξ_{jdt} that varies at the OD-day-time level and is unobserved by the econometrician. The coefficients α_i, β_i, and ρ_i represent the marginal (dis)utility of fare, crowding, and schedule deviation, respectively.

First consider the simple case where preference is homogeneous: $\alpha_i = \alpha$, $\beta_i = \beta$, and $\rho_i = \rho$. Normalizing the outside option utility to zero, the market share of time bin t in market (j, d) takes the standard logit form:

$$S_{jdt} = \frac{\exp(\delta_{jdt})}{1 + \sum_{\tau \in \mathcal{T}} \exp(\delta_{jd\tau})}, \tag{3}$$

where $\delta_{jdt} \equiv \alpha P_{jdt} + \beta \, \mathbb{E}_d[\text{Crowd}_{jdt}] + \rho \left| t + \mathbb{E}_d[\text{TT}_{jdt}] - t^{OA}_{jdt} \right| + \kappa_{jd} + \phi_t + \xi_{jdt}$ is the mean utility. After integrating over individuals, the rescheduling term in the aggregated equation uses t^{OA}_{jdt} rather than t^{OA}_{ijd}. The Control-as-Proxy imputation maps each rider to a percentile of the contemporaneous control distribution, so the representative imputed value at the product level varies across departure bins t. Following Berry (1994), the share equation can be inverted to yield the linear estimating

[1] The symmetry and linearity assumptions can be relaxed. We maintain them here because (1) the rescheduling cost is not the primary object of interest, and (2) the EBD incentivizes riders to reschedule to an departure earlier time.

equation:

$$\log S_{jdt} - \log S_{jd0} = \alpha P_{jdt} + \beta \widehat{\text{Crowd}}_{jdt} + \rho \left| t + \widehat{\text{TT}}_{jdt} - t^{OA}_{jdt} \right| + \kappa_{jd} + \phi_t + \xi_{jdt}, \tag{4}$$

where $\widehat{\text{Crowd}}_{jdt}$ and $\widehat{\text{TT}}_{jdt}$ are empirical counterparts for passenger expectations, which are proxied as fitted values from regressions of realized crowding and travel time on OD-day of week-time bin fixed effects. Because $\log S_{jd0}$ varies only at the market level, it is absorbed by κ_{jd}. The ratios β/α and ρ/α give the marginal willingness to pay (MWTP) to avoid one additional passenger per square meter for one minute of travel and the MWTP to reduce schedule deviation by one minute, respectively.

Passengers with different income levels may have different price elasticities and derive different levels of disutility from crowding and rescheduling. To capture this heterogeneity, we allow preferences to vary with income:

$$\alpha_i = \alpha_0 + \alpha_1 I_i, \qquad \beta_i = \beta_0 + \beta_1 I_i, \qquad \rho_i = \rho_0 + \rho_1 I_i, \tag{5}$$

where I_i is the passenger income, recentered at the sample mean. We simulate passenger income from a resident income distribution with an endogenous transportation mode choice. The construction of the income distribution of subway riders in each OD pair is described in Section 3.3. With heterogeneous preferences, the utility decomposes into a mean component δ_{jdt}, common to all passengers in the market, and an individual deviation μ_{ijdt} that depends on income:

$$\begin{aligned} V_{ijdt} = & \underbrace{\alpha_0 P_{jdt} + \beta_0 \widehat{\text{Crowd}}_{jdt} + \rho_0 \left| t + \widehat{\text{TT}}_{jdt} - t^{OA}_{jdt} \right| + \kappa_{jd} + \phi_t + \xi_{jdt}}_{\delta_{jdt}} \\ & + \underbrace{\alpha_1 I_i P_{jdt} + \beta_1 I_i \widehat{\text{Crowd}}_{jdt} + \rho_1 I_i \left| t + \widehat{\text{TT}}_{jdt} - t^{OA}_{jdt} \right|}_{\mu_{ijdt}}. \end{aligned} \tag{6}$$

The random-coefficient model no longer admits a closed-form expression for market shares. Predicted shares are computed by simulation over the income draws,

$$s_{jdt}(\delta, \theta) \;\approx\; \frac{1}{n_{jd}} \sum_{r=1}^{n_{jd}} \frac{\exp\left(\delta_{jdt} + \mu_{rjdt}\right)}{1 + \sum_{\tau \in T} \exp\left(\delta_{jd\tau} + \mu_{rjd\tau}\right)}, \tag{7}$$

where n_{jd} is the number of simulated passengers in market (j, d), drawn from the OD-specific income distribution. For a given vector of non-linear parameters $\theta = (\alpha_1, \beta_1, \rho_1)$, the mean-utility vector δ is recovered by the contraction mapping of Berry et al. (1995),

$$\delta^{h+1}_{jdt} = \delta^{h}_{jdt} + \log S_{jdt} - \log s_{jdt}\left(\delta^h, \theta\right), \quad h = 0, 1, \ldots, H, \tag{8}$$

and the structural error term is then

$$\xi_{jdt}(\theta) = \delta_{jdt}(\theta) - \left[\alpha_0 P_{jdt} + \beta_0 \widehat{\text{Crowd}}_{jdt} + \rho_0 \left| t + \widehat{\text{TT}}_{jdt} - t^{OA}_{jdt} \right| + \kappa_{jd} + \phi_t \right]. \tag{9}$$

Let Z_{jdt} be a vector of instrumental variables satisfying the orthogonality condition $\mathbb{E}[Z_{jdt}\, \xi_{jdt}(\theta_0)] = 0$. We estimate θ by GMM:

$$\hat{\theta} \;=\; \arg\min_{\theta} \; \big(Z'\xi(\theta)\big)^{\top} W \, \big(Z'\xi(\theta)\big), \tag{10}$$

where W is a positive-definite weighting matrix and $\xi(\theta)$ stacks the residuals across all markets and time bins.

2.2 Empirical Challenges

Estimating the above demand model poses two broad challenges: measurement and endogeneity. While detailed procedures of constructing those measurements and instrumental variables are described in Section 3.3, we provide brief intuition here.

2.2.1 Measurement

In-train crowding and the optimal arrival time are the two key explanatory variables in the model, but they are not directly observed in the data. The trip-level records from the automated fare collection (AFC) system contain only the origin and destination stations and the corresponding tap-in and tap-out timestamps. They do not reveal which route a passenger takes, which train she boards, or where she is located at any point during the trip. Measuring crowding requires assigning each passenger to a specific train at each point in time. We develop an iterative algorithm that jointly estimates the optimal route from the origin to the destination and the time spent walking, waiting, and transferring at each station, with in-train time imputed from the published timetable. Once every passenger's real-time location is determined, dividing the number of passengers sharing the same train segment by the available floor area yields the instantaneous crowding density in passengers per square meter. Summing this density over all segments of the trip gives the total in-train crowding Crowd_{jdt}, measured in passengers · minutes per square meter.

The optimal arrival time is also unobserved. Passengers choose departure times partly based on their preferred arrival times, and omitting this variable would confound the crowding coefficient: OD-time combinations with clustered optimal arrival times would exhibit both high crowding and high unobserved demand, biasing the estimate of β. We infer each passenger's optimal arrival time from "coworkers"—riders who arrive at the same destination around the same time but originate from stations unaffected by the early-bird discount. Because these coworkers face no fare incentive to shift their schedules, their arrival distribution reflects unconstrained preferences and serves as a natural proxy for the optimal arrival time.

Both measurement strategies exploit the *network structure* of the transit system. The train-assignment algorithm traces passengers across the subway network to compute crowding on shared segments. The coworker-based imputation identifies riders who converge at the same destination from different parts of the network. As we discuss next, the same network structure also underpins our instrumental variable strategy for crowding.

2.2.2 Endogeneity

In the demand model, both fare and crowding are potentially endogenous. The fare for a subway ride is a deterministic function of the track distance between the origin and the destination. Time variation in price comes from the early-bird discount, which offers a 30% reduction for trips originating from 16 designated stations before 7 AM. While the time cutoff of the EBD is arbitrarily set and plausibly exogenous to unobserved product-level demand shocks, two concerns remain. First, the EBD stations were not randomly selected; they are among the busiest suburban stations. We include the OD-day fixed effects κ_{jd} to absorb any time-invariant differences across stations, so α is identified solely from within-market time variation in fare. Second, the price variation introduced by the EBD is proportional to the fare *level*. If the sample only includes EBD stations, the time-bin fixed effects, ϕ_t, absorb all the variation in log price.[2] We address this by including a set of comparison stations on other suburban lines that exhibit similar ridership patterns but do not receive the early-bird discount. The comparison stations provide additional variation in the fare level that is not proportional to the EBD, strengthening identification of α.

The endogeneity of crowding poses a more fundamental challenge. Expected crowding is likely positively correlated with the unobserved demand shock ξ_{jdt}. A subway ride in a particular OD pair at a particular time is crowded not because passengers enjoy crowded trains, but because traveling in that OD pair at that time has high unobserved value *despite* the crowding. For instance, if many residents near the origin work near the destination and conventional work hours require arrival by 9 AM, commuting demand concentrates at specific departure times, generating both high crowding and a high ξ_{jdt}. This positive correlation biases the OLS estimate of β toward zero, or even flipping its sign to positive, understating the true disutility of crowding.

To break this endogeneity, we construct an instrumental variable based on *accidental companion passengers* (ACPs). An ACP is a rider who travels from a different origin to a different destination at a different time, yet has part of her trip overlapping with the focal passenger and shares the same subway car. The key assumption is that these ACP trips respond to unrelated demand shocks, so the number of ACPs shifts crowding on the relevant train segments without directly affecting demand for product (j, d, t).

The rescheduling term enters the demand equation without a separate instrument. Conditional

[2]The demand model includes price level instead of log price, and the model can be estimated with OD-day and time-bin fixed effects, using only EBD stations. But the model is only identified due to the specific functional form.

on κ_{jd} and ϕ_t, the residual variation in $|t+\widehat{\mathrm{TT}}_{jdt}-t^{OA}_{jdt}|$ reflects passenger-specific schedule constraints rather than route-level demand shocks: any common destination-side shock that drives both arrival timing and crowding is absorbed in κ_{jd}, and the within-cell variation that remains comes from the bin midpoint t and the predicted travel time $\widehat{\mathrm{TT}}_{jdt}$, neither of which is plausibly correlated with ξ_{jdt} given the fixed effects.

3 Background, Data, and Measurement

3.1 Background

Beijing is China's capital and second most populous city. Over the past two decades, it has built one of the world's largest subway systems. By 2016, the network consisted of 19 lines and more than 340 stations, carrying 3.7 billion annual passenger trips. The subway is especially important for long-distance commuting because of its speed and reliability. Excluding walking trips, the subway accounted for 15% of commuting trips and roughly 40% of commuting passenger mileage in 2014 (Beijing Transport Institute, 2015). A typical subway trip covers about 15 kilometers and takes approximately 37 minutes, including time spent accessing, waiting, and exiting.

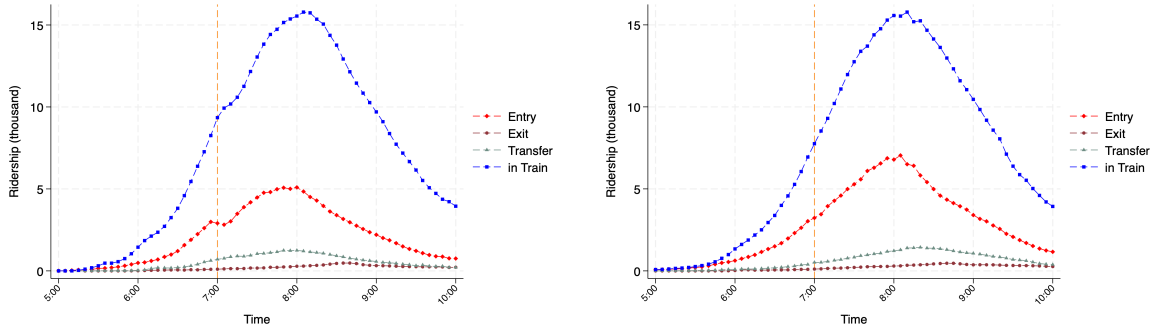
The fare for a subway ride is a step function of the track distance between the origin and the destination. A single trip starts at 3 RMB for distances under 6 kilometers, with the fare increasing by 1 RMB for each additional increment of up to 20 kilometers, eventually reaching 10 RMB for trips between 92 and 112 kilometers. Starting in December 2015, an early-bird discount (EBD) was introduced at 16 stations on two suburban lines, the Batong Line and the Changping Line. Passengers who tap in at one of these stations before 7 AM receive a 30% fare reduction.

The 16 EBD stations are among the busiest during the morning rush, but they are not the only busy stations. Other suburban lines connecting dense residential neighborhoods with the city center exhibit similar ridership patterns. The EBD was later expanded to eight additional stations on Line 6 by the end of 2016, which falls outside our sample period. We select 16 stations on Lines 5 and 6 as comparison stations. Line 6 runs parallel to the Batong Line, and the eight stations we select received the EBD after our sample period. Line 5 cuts through the city center from north to south, with its northern branch paralleling the Changping Line and passing through dense residential neighborhoods. The eight Line 5 stations we choose have been in active discussion for inclusion in the EBD program. Appendix Figure B.1 shows the location of EBD and control stations in the Beijing Subway network.

Figure 1 compares passenger flows at EBD and control stations during morning hours. The two groups exhibit strikingly similar in-train ridership profiles, with both peaking at approximately 15,000 to 16,000 passengers around 8:00 to 8:30 AM. The similarity supports the use of control stations as a valid comparison group. The figure also reveals the behavioral response to the EBD. At

EBD stations, the entry flow displays a visible dent just after 7 AM, with a corresponding bunching of entries just before the cutoff. This pattern is consistent with riders advancing their departures to qualify for the 30% discount. No such discontinuity appears at control stations, where the entry flow rises smoothly through the morning peak. The in-train passenger counts, which cumulate entries from upstream stations, show no comparable discontinuity in either group, confirming that the EBD-induced bunching is a localized phenomenon at the station entry level.

Figure 1: Passenger Flows during Morning Hours



(a) EBD Stations

(b) Control Stations

Notes: This figure plots the number of passengers (in thousands) at EBD stations (panel a) and control stations (panel b) by 5-min time bin during morning hours. The four series represent passengers entering the station, exiting the station, transferring through the station, and riding in trains passing through the station. The dashed vertical line marks the 7 AM EBD cutoff. Both groups of stations exhibit similar in-train ridership levels, peaking around 8:00 to 8:30 AM. At EBD stations, entry passengers display a visible dent just after 7 AM as some riders advance their departures to qualify for the discount, with a corresponding bunching just before 7 AM. No such discontinuity appears at control stations.

3.2 Data

The primary data source is trip-level records from the Beijing Subway's automated fare collection (AFC) system (Beijing Municipal Institute of City Planning and Design, 2015). The Beijing Subway uses an electronic smartcard as its payment method. The smartcard is widely popular because it is easy to use and only card users qualify for fare discounts. By comparing official ridership statistics with our data, we estimate that between 90% and 95% of subway trips were paid by smartcard around the time of the EBD. Each record captures the tap-in and tap-out stations and timestamps for a single trip, and trips can be linked across days through an anonymized card identifier. The main data used in this paper consist of the universe of smartcard trips in 21 working days in 5 non-consecutive weeks in 2016.

We supplement the AFC records with several additional data sources. We geocode all subway stations and collect the track distance and fare for each station pair from Beijing Subway's website. We collect train frequencies by line and station, as well as the types of passenger cars and the number

of cars per train, from which we compute the total floor area of each train. For train floor area, we do not distinguish between seating areas and standing areas. We also use the 2015 Beijing Household Travel Survey (Beijing Transport Institute, 2015), a large representative survey of approximately 100,000 households that records each respondent's home and work locations, daily travel details, and transportation mode choices. We further use grid-level commuting flow data from Baidu Maps to construct OD-pair-specific income distributions; the data and our matching procedure are described in Section 3.3.4. Appendix A provides a detailed description of all data sources used in this paper.

3.3 Measurement

3.3.1 Optimal Route, Train Assignment, and Crowding

To measure in-train crowding, we need to determine each passenger's location at every point during her trip. There are two key challenges to this task. First, the AFC data record only each passenger's tap-in and tap-out stations and times, not which trains she boards. We also do not observe trains' real-time locations or the crowding density on platforms or inside cars.[3] Second, with 19 lines and more than 340 stations, the Beijing Subway is a complex network, and many OD pairs admit multiple feasible routes. We need to determine which route each passenger takes.

Optimal route and train assignment. 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,

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

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 $\mathcal{P}_{jdt}$. We observe entry time (t^O) and exit time (t^D), but not the transfer stations or time arriving at and departing from those stations (t^1, t^2, etc).

The total travel time along route $\mathcal{P}$ can be decomposed as

$$TT_{jdt}^{\mathcal{P}} = \underbrace{ST_{jdt}^{\mathcal{P}}}_{\text{in-train}} + \underbrace{AT_{t^O}^{O}}_{\text{entry}} + \underbrace{\sum_{R \in \mathcal{R}_{\mathcal{P}}} RT_{t^R}^{R}}_{\text{transfer}} + \underbrace{ET_{t^D}^{D}}_{\text{exit}}, \tag{11}$$

where $ST_{jdt}^{\mathcal{P}}$ is the in-train time, defined as the sum of scheduled train travel times between adjacent stations along the route; $AT_{t^O}^{O}$ is the entry time from tapping in to boarding the first train; $ET_{t^D}^{D}$

[3] Most trains in the Beijing Subway are equipped with real-time location devices, which are used to display estimated arrival times at stations. Platforms and passenger cars also have cameras that could potentially be used to derive crowding measures. However, we do not have access to those data.

is the exit time from alighting to tapping out; and $RT^R_{t^R}$ is the transfer time at transfer station R when arriving at time t^R. Of these, only the total travel time $TT^{\mathcal{P}}_{jdt}$ (observed from the tap-in and tap-out timestamps) and the in-train time $ST^{\mathcal{P}}_{jdt}$ (computed from the published timetable under the assumption that trains always operate on schedule) are directly known.

We estimate the entry, exit, and transfer times through an iterative calibration procedure applied to the entire subway network. We begin by assigning initial guesses for each in-station component. These initial values allow us to back out arrival times at transfer stations and calculate the total travel time for each candidate route. Using the calculated travel times, we apply Dijkstra's algorithm to identify the optimal route for each OD pair.

We update the estimates of in-station time components by regressing the observed median total travel time $\hat{TT}_{jdt^O}$ for trips in each OD-tap-in-time bundle on the timetable-based in-train time $\hat{ST}_{jdt^O}$ and a set of station-by-purpose (entry, exit, transfer) specific fixed effects capturing in-station time cost:

$$\hat{TT}_{jdt^O} = \hat{ST}_{jdt^O} + AT^O_{t^O} + \sum_{R \in \mathcal{R}} RT^R_{t^R} + ET^D_{t^D} + \varepsilon_{jdt^O}. \tag{12}$$

The regression is estimated by OLS, pooling across the entire subway network and weighting by the number of trips in each jdt^O. There are enough degrees of freedom to flexibly identify entry, exit, and transfer times specific to each station. We recalculate the optimal route and in-train time using the updated estimates, and iterate until all in-station time parameters and the implied route choices converge.[4]

Given a trip's optimal route, Equation 12 faces a collinearity problem. The total travel time is observed, the in-train time is determined by the timetable, and the error is mean-zero, so the in-station time components must sum to a constant equal to the mean total travel time minus the in-train time. To break the collinearity, we normalize the mean exit time to two minutes.

Appendix Table B.1 summarizes the estimated in-station times. Entry time, from tapping in at the origin station to boarding the train, averages 3.5 minutes with a standard deviation of 1.84 minutes. This variation may reflect station configurations and train frequencies. Exit time averages two minutes by construction. Transfer time, at an average of 4.43 minutes, is the longest in-station component, with some transfer stations requiring more than six minutes. The last column of the table reports the number of observations used to estimate each component. Because all trips have an entry and exit station, these two components are estimated from approximately 33.4 million trips. The average trip involves slightly more than one transfer, so transfer time is estimated from 34.8 million trips.

Not all passengers traveling from O to D with tap-in time t^O arrive at the same time. There is a

[4]While the optimal route may in principle vary over the course of the day due to time-varying in-station congestion, in practice the fastest route for a given OD pair rarely changes by time of day. This is because the second-best routes typically cost substantially more time. We therefore assign the most common optimal route across all departure times to each OD pair.

distribution of total travel time, which we attribute to idiosyncratic factors in the in-station segments. For example, some passengers walk faster than others, some are lucky to catch a departing train while others must wait for the next one, and some may stop at a convenience store or vending machine inside the fare gates. While assuming all passengers take the same optimal route, we treat the sources of travel time variation as agnostic and proportionately scale each in-station segment of the median passenger's time cost to fit the observed total time for each individual trip. For example, suppose the imputed travel time for a trip in jdt^O is 40 minutes, composed of 10 minutes entry, 5 minutes transfer, 20 minutes in-train, and 5 minutes exit. If an actual trip in that bundle takes 44 minutes (10% longer than imputed), we assign 12 minutes for entry, 6 minutes for transfer, and 6 minutes for exit, while the in-train time remains fixed at 20 minutes.

The route assignment algorithm rests on several assumptions that merit discussion. First, train speed is assumed not to vary with crowding. During the study period, the Beijing Subway was largely new and equipped with advanced computerized control systems. Stopping times at stations are automated, and most stations have platform screen doors to prevent accidents. Unlike some older systems, delays were rare. We show in Section 4 that the empirical relationship between crowding and travel time is negligible. Second, we assume that passengers do not detour to avoid crowded routes. Detours typically cost substantially longer time, and Section 4 provides evidence that route-switching to avoid crowding is uncommon. Third, we assume that the distribution of travel time within an OD-time bundle reflects idiosyncratic shocks and is unrelated with crowding levels.

With each passenger's route and time decomposition in hand, we can determine her real-time location at any point during the trip. We define a *segment* s as the track between two adjacent stations. Each passenger's trajectory is discretized into segment-by-five-minute-bin cells. When a passenger traverses multiple segments within a single five-minute bin, her presence is distributed across segments in proportion to the time spent on each. When a single segment spans multiple five-minute bins, it is partitioned into sub-segments assigned to the corresponding bins. Summing over all passengers yields $N_{s\tau}$, which is the weighted ridership on segment s during time bin τ.

Crowding. We focus on in-train crowding.[5] We define the *in-train crowding density* $k_{s\tau}$ as the number of passengers per square meter on segment s in time bin τ, obtained by dividing $N_{s\tau}$ by the total train floor area available during the time bin, accounting for the floor area of each train and train frequency.[6] $k_{s\tau}$ reflects the degree of crowding a passenger experiences at a given time and has

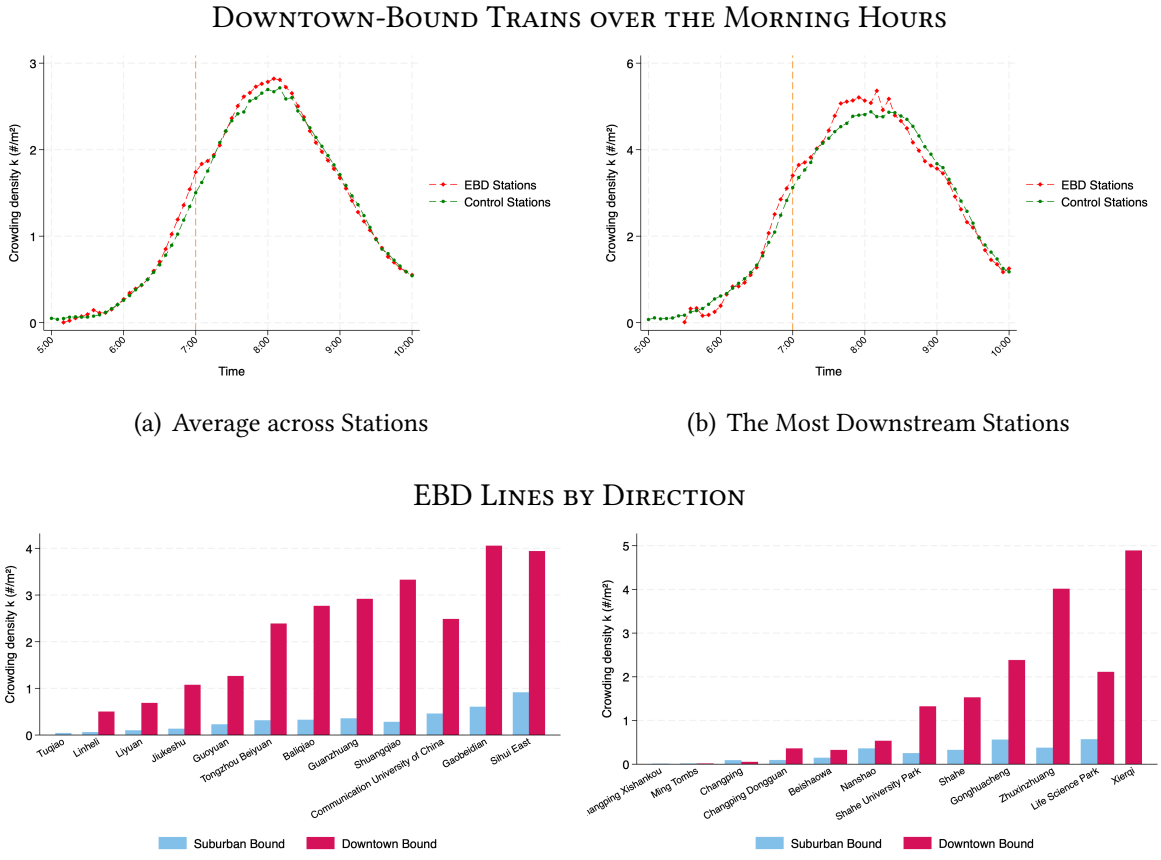
[5]Passengers may also derive negative utility from crowded stations. Our route assignment algorithm also allows us to calculate the number of passengers in each station at any given time. We focus on in-train crowding because it is arguably the most important dimension of the disamenity. Including in-station crowding is possible, but because the station is much larger and passenger distribution is uneven, relevant crowding density is difficult to impute. We show in robustness checks that including in-station crowding does not affect the estimates of the WTP to avoid in-train crowding.

[6]Each train in our sample consists of six passenger cars measuring 19 meters long and 2.8 meters wide, providing a total floor area of approximately 320 square meters. We do not subtract floor space occupied by seats and other on-train equipment, nor do we model the probability of finding a seat or differential utilities from sitting and standing. Using

units of persons per square meter.

For each passenger i, the *total in-train crowding experience* K_i aggregates the crowding density over all segment-time bins along her route, weighted by the time spent in each. The in-train crowding experience has units of person-minutes per square meter, capturing the total discomfort as the product of the intensity and the duration of crowding exposure. K_i is the main crowding variable in the demand model. We aggregate K_i to the product level by computing the mean across all trips within each OD-day-15-minute-bin (jdt) cell.

Figure 2: In-Train Crowding Density



(a) Average across Stations

(b) The Most Downstream Stations

(c) Batong Line

(d) Changping Line

Notes: This figure depicts in-train crowding density k (persons/m^2). The two graphs in the top row show downtown-bound trains over the course of the morning peak hours. Panel (a) averages across stations on EBD and control lines; panel (b) shows the most downstream station from each line. The two graphs in the bottom row show crowding density by station and direction for the two EBD lines during morning peak hours (6:30–8:30 AM). Stations are listed from the most upstream on the left to the most downstream on the right. Downtown-bound trains are substantially more crowded than suburban-bound trains. Crowding is computed using the full universe of rides.

Because the stations in our sample sit at the far ends of suburban lines, the vast majority of morning passengers travel toward downtown. The top row of Figure 2 show the average in-train crowding

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. Train frequency is from the train timetable.

density during morning hours in downtown-bound trains passing through EBD and control stations. The two groups exhibit strikingly similar levels and time profiles. Crowding density averages about 1.5 persons per square meter around 7 AM, peaks at roughly 2.8 persons per square meter near 8:15 AM, and declines to 0.5 persons per square meter by 10 AM. The EBD and control series nearly overlay except in a narrow window around the 7 AM cutoff, where bunching at EBD stations creates a visible hump. The right sub-panel narrows the comparison to the most downstream station on each line, where trains have accumulated passengers from all upstream stations. Although this sub-panel draws from only four stations (one from each line), the crowding patterns remain closely aligned between EBD and control lines. With passengers boarding from upstream stations, trains passing the most downstream stations are substantially more crowded. At around 8:20 AM, crowding density at the most downstream EBD stations peaks at approximately five persons per square meter.

The two graphs in the bottom row of Figure 2 show that downtown-bound trains are far more crowded than suburban-bound trains on the two EBD lines. Trips in both directions qualify for the EBD and are included in our sample. The substantial difference in crowding between directions provides variation that helps identify potentially nonlinear crowding disutility at different density levels.

Passengers form expectations about crowding based on their commuting experience. To proxy for these expectations, we regress realized crowding on OD-by-day-of-week-by-time-bin fixed effects and use the fitted values $\widehat{\text{Crowd}}_{jt}$ as the expected crowding measure in the demand model. These fixed effects capture the systematic, predictable component of crowding patterns, precisely the information available to regular commuters. The same approach is applied to construct the expected travel time. The R^2 from this regression exceeds 0.98, indicating that passengers' crowding experiences are highly predictable from the OD pair, day of week, and time of day.

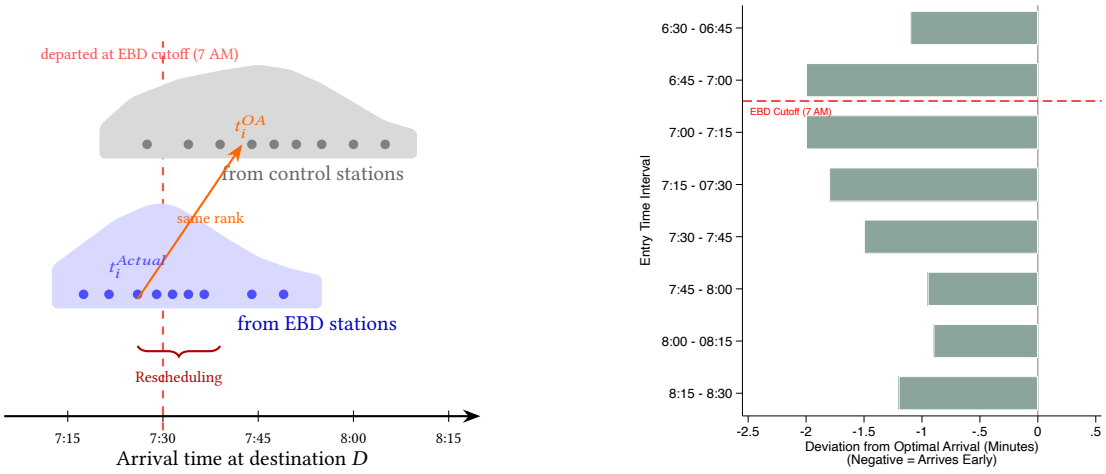
3.3.2 Optimal Arrival Time and Rescheduling

The rescheduling cost in the demand model requires measuring the optimal arrival time t_{ijd}^{OA} for each passenger. This variable is fundamentally unobserved. Observed arrival times may not be optimal because passengers trade off schedule deviation against fare savings and expected crowding. In particular, the EBD incentivizes passengers to depart earlier than they otherwise would, so observed arrival times of trips from EBD stations are also likely to be earlier than optimal. Exploiting the network structure of the subway system, we construct the optimal arrival time using trips that arrive at the same destination station but originated from non-EBD stations.

The key idea is that passengers arriving at the same destination in the morning peak hours as the focal passenger are likely responding to the same need for travel. For example, they work for the same company and need to show up for work on time. We thus call these passengers "workers." However, coworkers originating from control stations face no fare incentive to shift their departure times, so their arrival time distribution reflects unconstrained scheduling preferences. If control

station riders travel to the same destination on the same day as EBD riders, the two groups face similar workplace constraints and destination-side demand conditions. The control riders' arrival distribution therefore serves as a natural counterfactual for what EBD riders would have chosen absent the discount.

Figure 3: Optimal Arrival Time Imputation and Validation



(a) Quantile Matching Procedure

(b) Mean Schedule Deviation by Entry Time

Notes: Panel (a) illustrates the quantile matching procedure. The blue distribution shows observed arrivals at destination D from EBD stations, shifted earlier by the fare discount. The gray distribution shows arrivals from control stations traveling to the same destination on the same day. Each EBD rider is matched to the control distribution at the same within-group percentile rank, yielding the imputed optimal arrival time t_i^{OA}. Panel (b) plots the mean signed deviation (actual minus optimal arrival time, in minutes) for EBD station trips by 15-minute entry time bin. Negative values indicate that passengers arrived earlier than the imputed optimal time. The dashed line marks the EBD discount cutoff at 7 AM.

Figure 3 panel (a) provides a graphical illustration of this approach. For a given date d and destination station D, we collect all trips arriving at D during the morning rush from control stations. This set of trips forms the control arrival distribution. We then collect the corresponding trips from EBD stations. Because the EBD discount shifts some passengers to earlier departures, the EBD arrival distribution is likely shifted to the left relative to the control distribution, as illustrated in Figure 3 panel (a). We perform quantile matching to map each EBD rider to the control distribution. Specifically, let p_i denote the percentile rank of EBD rider i within the EBD arrival distribution for that date-destination cell. The imputed optimal arrival time is

$$t_i^{OA} = F_{\text{Ctrl}}^{-1}(p_i), \tag{13}$$

where F_{Ctrl}^{-1} is the inverse cumulative distribution function of the control group's arrival times. The matching is performed at high precision using continuous seconds. Cells with fewer than ten EBD trips or fewer than ten control trips are excluded from the procedure.

For trips originating from control stations, we set the optimal arrival time equal to the actual arrival time, $t^{OA}_{\text{Ctrl}} = t^{Actual}_{\text{Ctrl}}$, so that rescheduling cost is zero by construction. This ensures that any nonzero rescheduling cost in the estimation comes from the EBD group, where the fare discount induces measurable departures from schedules not directly unaffected by the EBD.

This approach has several advantages. It relies entirely on contemporaneous data, eliminating concerns about temporal shifts in commuting patterns that would arise from using pre-policy historical records. By matching within the same date-destination cell, it captures day-specific demand conditions such as weather or local events that affect both EBD and control riders symmetrically. This approach is essentially non-parametric. It assigns coworkers through matching two *distributions*, rather than directly matching individuals. Without knowing the exact workplace for each rider, assigning coworkers to each rider would require more assumptions and likely introduces more biases.

A key assumption underlying the method is that passengers from EBD and control stations traveling to the same destination face similar optimal arrival incentives. Appendix Figure B.2 provides supporting evidence by comparing the destination distributions of the two groups. Of the 241 unique destination stations in the sample, 188 are served by riders from both EBD and control origins. The remaining 53 single-group destinations account for a negligible share of total ridership (0.55% for EBD-only destinations and 1.21% for control-only destinations). The Pearson correlation between the two destination share vectors is 0.46, and the Spearman rank correlation is 0.68, indicating that the two groups share broadly similar destination patterns despite originating from geographically distinct lines.

Figure 3 panel (b) validates the imputed rescheduling measure by plotting the mean signed deviation (actual minus optimal arrival time) for trips originated from EBD stations across 15-minute entry time bins. Negative signed deviations indicate that the passenger arrived earlier than optimal, which is the expected direction for EBD riders who advance their departures to qualify for the discount. Riders who tap in before the 7 AM cutoff arrive approximately one to two minutes earlier than their imputed optimal time. After 7 AM, when the discount no longer applies, deviations become smaller. The largest deviations are observed for trips that depart around the cutoff. The clear discontinuity at the cutoff confirms that the imputed optimal arrival time captures meaningful variation in rescheduling behavior. The vast majority of the deviations calculated for EBD trips are negative. The rescheduling variable used in the demand model is the absolute deviation between the expected arrival time and the imputed optimal arrival time, $|t + \widehat{\text{TT}}_{jdt} - t^{OA}_{jdt}|$.

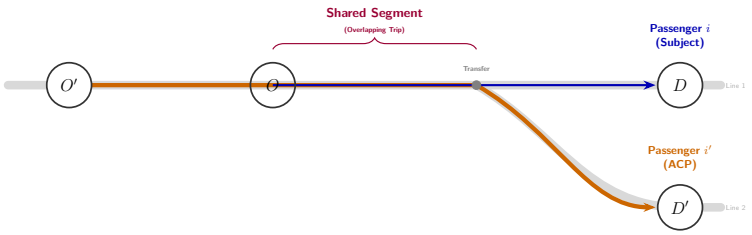
3.3.3 Accidental Companion Passengers (ACP)

Crowding is endogenous. It reflects the number of passengers on the same trains and is positively correlated with unobservable demand shocks ξ_{jdt}. This correlation biases the crowding disutility coefficient toward zero or even positive values, making crowding appear less costly than it truly

is. A valid instrumental variable needs to be unrelated to the demand shocks, yet remains strongly correlated with crowding.

The common demand shocks that prompt passengers to take the same train may come from three distinct sources: at the same *origin*, at the same *destination*, and for the same *purpose*. For example, commuters may board the same train at the same time because it is the most convenient departure time while alternative transportation modes are limited near where they live (common demand shock at O); downtown-bound trains are crowded because economic and social activities are concentrated in the city center (common demand shock at D); and peak-hour trains are crowded because people commute to work (demand shock due to common purpose). We break those sources of endogeneity while remaining relevance by constructing an instrumental variable based on "accidental companion" trips: trips travel from a different origin, to a different destination, and taken by a different type passengers, yet overlap with the focal trip in both space and time due to the configuration of the subway network.

Figure 4: Accidental Companion Passengers



Notes: Passenger i travels from O to D. Passengers i' and i'' board upstream of O and share part of i's train segment. i' alights at D' (far from D) and qualifies as an ACP; i'' alights at D'' (near D) and is excluded due to potential spatially correlated demand shocks.

Figure 4 illustrates the construction. For passenger i's trip in (OD, d, 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. Because the ACP trip overlaps with (OD, d, t) but has a different origin and destination, its starting and ending times also must be different.

Our main analysis focuses on frequent subway users, so our preferred definition of ACPs are infrequent users, defined as those who take no more than three trips per week. Frequent users likely ride subway for commute, while infrequent users ride the subway for different purposes. Moreover, one may worry that frequent riders may strategically adjust their travel timing in response to crowding conditions, creating a negative correlation between ACP-based measures and unobservable demand shocks that would bias the IV estimates downward. Infrequent riders, by contrast, are less likely to monitor and respond to day-to-day fluctuations in crowding, making their presence on a given train segment more plausibly exogenous.[7] To further mitigate the concern of endoge-

[7]To see why including frequent riders may bias estimates downward, note that frequent ACPs may reduce their like-

nous responses to expected crowding, we further restrict ACPs to infrequent users boarding from an upper-stream station of O.

For the instrumental variable to be valid, we are essentially assuming that infrequent subway users do not form an expectation of the day-to-day variation in the in-train crowdedness downstream. This assumption is plausible because passengers who ride the subway only occasionally lack the necessary knowledge for the prediction. Notice that we do not require infrequent subway riders to know nothing about the crowding status. All our models control for OD-by-date fixed effects and time bin fixed effects, so we allow ACPs to form expectation based the specifics of the route and the day (weather, traffic on the route, etc), and the typically crowdedness at any given time bin in the subway system.[8]

Between correlated demand shocks at the origin and those at the destination, we hypothesize that correlated shocks at the destination is a more important concern, as morning rush hour trips are probably responding to a common check-in time in employers near an exit station. In the empirical analysis, we vary the degree of stringency in selection ACPs with O', D' to test the possible sources of endogenous demand shocks.

We define *accidental companion time* (ACT) as the weighted sum of accidental companion passengers (ACPs), where the weight is the shared travel time. Let $\mathcal{I}'$ be the set of ACPs. The ACT for passenger i is then given by:

$$ACT_{ijdt} = \sum_{i' \in \mathcal{I}'} m^{i,i'}. \tag{14}$$

ACT weights each companion passenger by the duration of shared travel, so that a co-rider who overlaps with i for many segments contributes more than one who shares only a brief portion of the route. Compared with ACP, ACT is likely more relevant and contributes to a stronger first stage. Finally, because expected crowding is based on historical crowding levels, the ACPs and ACT are constructed using trips from the same historical dates used to form crowding expectations.

The ACT instrument is structurally analogous to the IV strategy in Hortaçsu et al. (2024), who instrument for endogenous airline fares using the opportunity cost of seat capacity driven by demand from other itineraries sharing the same flight leg. In both settings, the identifying variation comes from travelers who share the same physical infrastructure but serve distinct markets. The exclusion restriction requires that demand from these companion travelers affects the focal outcome only through the endogenous variable. Our spatial and frequency restrictions on ACPs ensure that companion demand is orthogonal to the focal market's unobserved demand. While the validity of the instrument is not directly testable, we provide evidence consistent with the underlying validity

lihood of riding when downstream demand is high to avoid crowding. This strategic response generates $\text{Cov}(ACT_{jt}, \xi_{jdt} \mid X_{jdt}) < 0$, which, given the positive first-stage relationship between ACT and crowding, pulls the GMM estimate below the true parameter. Section 5 presents robustness results under varying frequency thresholds and distance restrictions.

[8] In some specification, we control for fixed effects of time bin-by-line of the destination station, which further allows ACPs to know the time-specific conditions at their destination lines.

assumption.

3.3.4 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 combining outside data sources.

We use grid-level commuting flow 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. It is able to track millions of smart mobile devices (mostly smartphones) with precise and high-frequency location data. In particular, it can determine a device's usual nighttime location (which it refers to as home) and its usual daytime location (work). The commuting flow data contain the number of commuters between all pairs of home and work locations in Beijing, where each location is a small grid with an edge of approximately 700 meters (and an area of about 0.5 square kilometers). The data are based on the usual location patterns in the three months ending in November 2021 and cover 7.6 million unique devices with home locations distributed across roughly 15,000 grids.[9] Importantly, the data also report the share of commuters in each home-work location pair in each of five monthly income brackets: 0 to 2,499 RMB, 2,500 to 3,999 RMB, 4,000 to 7,999 RMB, 8,000 to 19,999 RMB, and 20,000 RMB and above. 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 includes the make and model of the device, applications installed on the device, some disclosed information to Baidu and its partnering applications, and detailed location information. We do not have access to those underlying user characteristics, the exact algorithm Baidu uses, or the training data it uses to validate the imputations.

We assign commuters from the Baidu data to subway station pairs. We define a station's catchment as an area within a three-kilometer radius. Data from the Beijing Household Travel Survey (BHTS) (Beijing Transport Institute, 2015) show that very few people choose to ride the subway if they live or work more than three kilometers away in straight-line distance from a station. Catchment areas from different stations can overlap. Users in multiple catchment areas are probabilistically assigned to stations 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 commuters in the Baidu data, we define the user pool of each station pair OD.

[9] Beijing has approximately 22 million residents and a land area of 16,000 square kilometers. The grids do not cover the mountainous regions where few people live and bodies of water.

For each OD pair, we estimate a continuous income distribution from the grouped income data by fitting a beta distribution via grouped maximum likelihood estimation. We assume the OD-specific income distribution is time-invariant. For the random-coefficient demand model, we draw 150 simulated passenger incomes from the fitted distribution for each OD-date market. Every OD pair in the working sample is successfully assigned an income distribution, so no markets are dropped at this step.

Figure B.3 plots the distribution of simulated monthly incomes. The weighted mean income is approximately 8,770 RMB per month. The median monthly income of potential subway riders is about 8,000 RMB, and the 90th percentile is around 18,000 RMB.

Figure 5 displays the geographic distribution of mean monthly income (in thousand RMB, adjusted to 2016 levels) by residential location (Panel a) and workplace location (Panel b). Each cell aggregates a 2×2 block of the underlying Baidu grid, giving cells approximately 1.4 km on a side. Within each cell, we compute the mean monthly income as the population-weighted average across the five Baidu income brackets. The 2016 subway network is overlaid in red and the Beijing municipality boundary in dark grey. The two panels show complementary spatial structure. Residential income is highest in the northwest (Haidian) and northeast (Chaoyang/CBD periphery), and high-income residents are spread across suburban areas. Workplace income, by contrast, concentrates much more tightly along the central business corridors (Guomao/CBD, Financial Street, Zhongguancun). This residential–workplace divergence is the geographic underpinning of the morning peak commute we study: many high-income suburban residents commute downtown in the morning and reverse the trip in the evening. The EBD stations sit precisely on the suburban end of this commute, capturing a large share of the relevant origin-side flow.

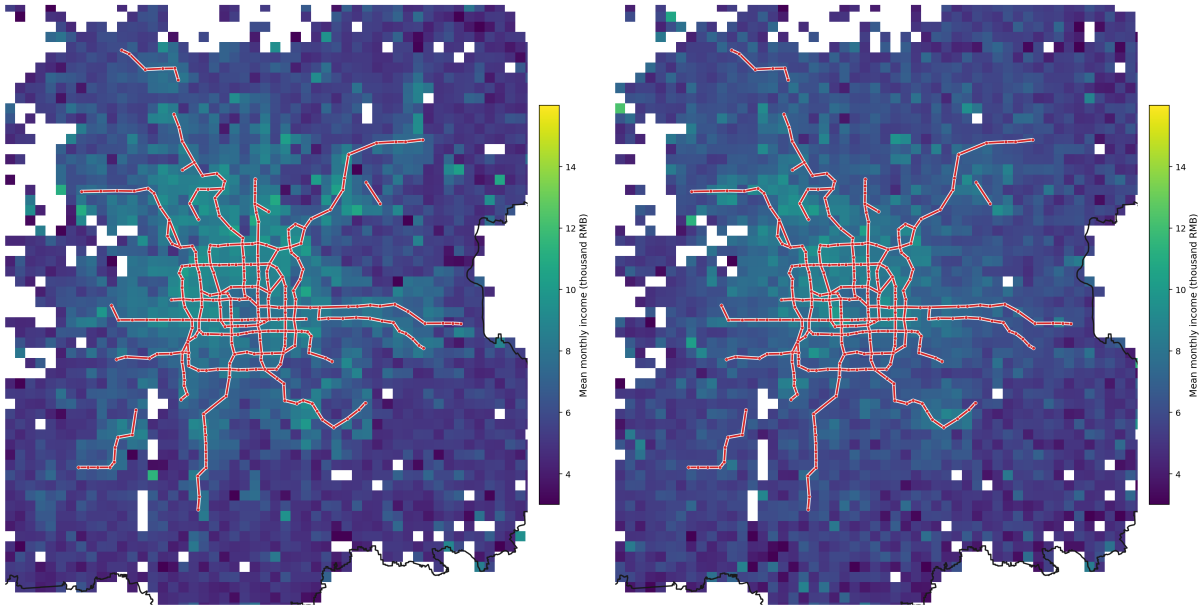
3.3.5 Transportation Mode Choice

The Baidu commuting flow data capture trips in all modes of transportation. To define market sizes and the outside option in the demand model, we need to predict which commuters use the subway. We fit a multinomial logit model of mode choice using the 2015 BHTS with predictors that are also available in the Baidu data, namely income levels and origin-destination locations.

The BHTS reports only household income (in five bins), not individual income. We predict individual income by regressing household income (converted to log RMB) on flexible functions of each working household member's gender, age, education level, industry and occupation, their interactions, and a household fixed effect. The estimated coefficients are used to predict individual income subject to the constraint that all working members' predicted incomes sum to the observed household total.

We then fit a multinomial logit where each observation is a commuting trip. The dependent variable is the main transportation mode, defined as the mode used for the longest leg. Modes include the subway, bus, and car, with other modes (bicycles, electric scooters, walking) as the other cate-

Figure 5: Mean Monthly Income by Cell



(a) Residential cell (b) Workplace cell

Notes: Cell-level mean monthly income (thousand RMB, adjusted to 2016 levels) computed from the Baidu commuting portrait data as the population-weighted average across income brackets, with bracket midpoints as the per-bracket value. Cells are 2×2 aggregations of the underlying Baidu grid (cells of approximately 1.4 km on a side); cells with fewer than 5 commuters are dropped. The 2016 subway network (lines opened by 31 December 2016) is overlaid in red; the Beijing municipality boundary in dark grey.

gory. Explanatory variables include individual income, its square, and trip distance. Table 1 reports the estimation results. Higher income is positively associated with all motorized modes relative to walking, biking, or scooting, with the largest effects for subway and car. Trip distance is also strongly positively associated with motorized modes.

Figure 6 shows mode choice shares from the BHTS by income and trip distance during the morning peak (6 to 8 AM). The subway share rises sharply with trip distance, especially for middle- and high-income commuters. For commuters with an income above 8,000 RMB (the sample median), more than 40% choose subway for commutes longer than 12 km. In contrast, very few choose to ride the subway for trips under 5 km.

Bus is primarily used by low-income individuals for middle-to-long distance trips. Car uses are concentrated for the high. Car use increases with both income and distance, exceeding 30% for high-income commuters on longer trips. But interestingly, subway edges out in very long trips for even the highest-income individuals. Other modes, primarily walking and cycling, dominate only for short trips (under 2 km) and decline rapidly as distance grows, regardless of income.

We use the fitted mode choice model to define market sizes. We predict user mode choices from the Baidu data using predicted mode shares based on income and travel distance. Let $\hat{s}_j^{\text{sub}}$ denote the

Table 1: 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	

Notes: Multinomial logit estimates. The baseline category is "other" modes (bicycles, electric scooters, walking). Standard errors in parentheses.

subway mode share for OD pair j derived from the BHTS. The total market size is $M_{jd} = Q_{jd}^{\text{sub}} / \hat{s}_j^{\text{sub}}$, where Q_{jd}^{sub} is the observed subway ridership from the smartcard data. The outside good encompasses all non-subway modes and is given by $Q_{jd}^{\text{out}} = M_{jd} - Q_{jd}^{\text{sub}}$.

3.4 Sample and Summary Statistics

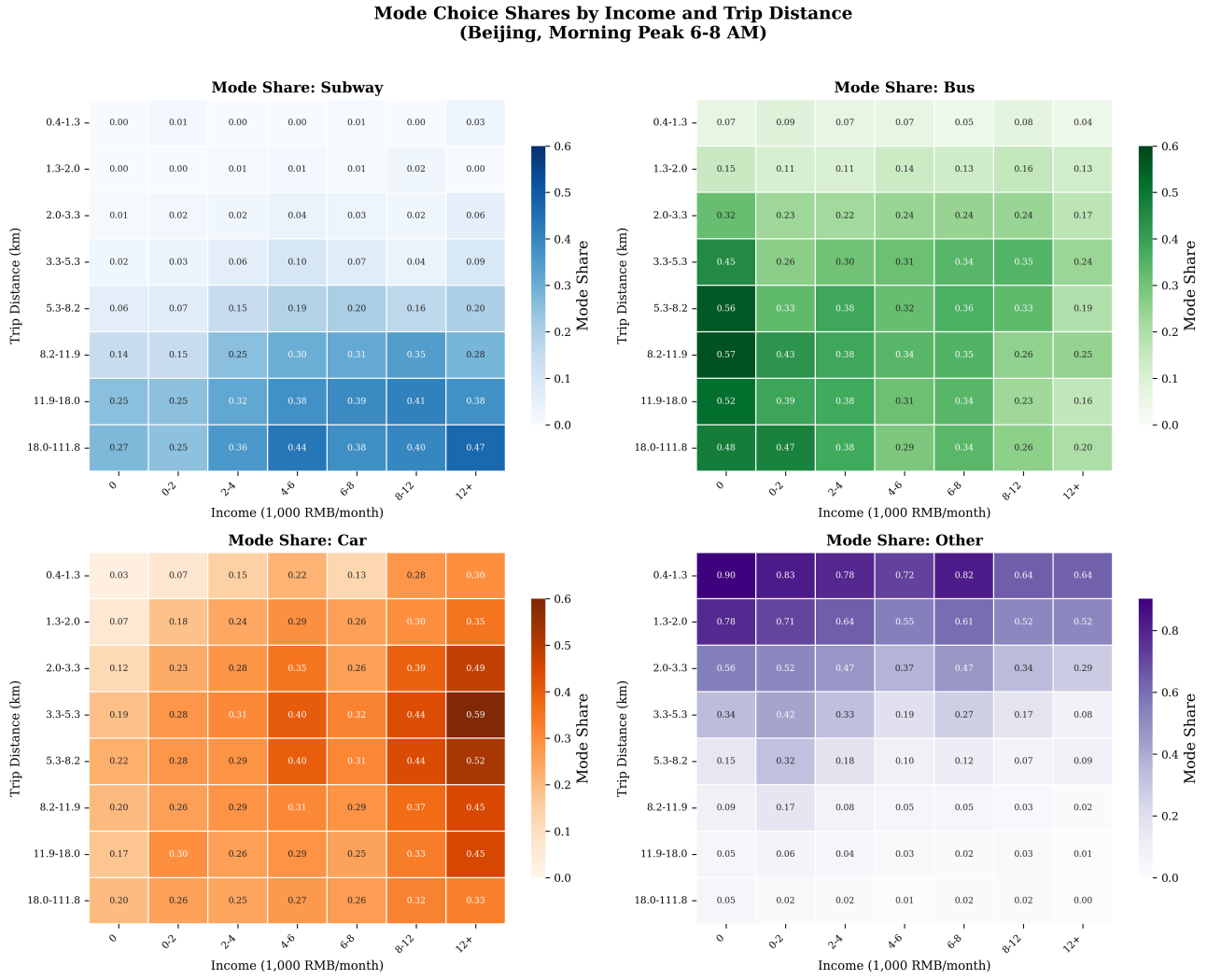
Our sample includes trips originating from the 16 EBD stations and 16 control stations and ending at any of the 342 stations in operation. We restrict to trips with a tap-in time between 6:30 and 8:30 AM. The EBD discount applies to passengers who tap in before 7 AM, so the two-hour window spans across the time cutoff. While the first train departs the terminal station around 5 AM on all four lines, ridership is low in the early hours (Figure 1), and we begin the sample at 6:30 AM.

The sample covers five non-consecutive weeks spanning March, June, September, November, and December of 2016. Weekends and public holidays are excluded because the EBD does not apply on those days and because travel patterns differ substantially between workdays and non-workdays. This leaves 21 working days in the sample.

We define a *market* as an OD-by-date pair within the selected time window. In each market, passengers choose when to tap into the origin station. Each *product* is a 15-minute departure time bin, yielding eight products per market. The outside good represents traveling between the same origin and destination during the morning rush hours by alternative modes. Our baseline regressions focus on frequent riders, defined as passengers who average at least three subway trips in that week. The crowding measure, however, is always computed using the full universe of rides to accurately capture system congestion.

Many markets are small. We focus on large markets by including those that have all eight products with non-zero trips. The final sample contains approximately 3.8 million trips across 40,191

Figure 6: Mode Choice Shares by Income and Trip Distance



Notes: Mode shares across income and trip distance bins from the BHTS 2015 survey (morning peak, 6 to 8 AM). Each cell shows the fraction of travelers in that income-distance bin choosing the given mode. Income is monthly household income per capita (1,000 RMB). Distance bins are based on quantiles of reported trip distance.

markets. The EBD subsample includes 17,777 markets covering 1,621 unique OD pairs; the control subsample includes 22,414 markets covering 1,920 unique OD pairs.

Table 2 reports summary statistics for the key variables in the treatment and control groups. The two groups are broadly comparable. The average number of subway passengers per product is 11.0 in the EBD sample and 12.4 in the control sample. Trips from both groups are long, averaging 20.0 km and 46.9 minutes for EBD stations and 18.8 km and 41.6 minutes for control stations. This is not surprising because both EBD and control stations are at the far ends of suburban lines. The average in-train time is 32.0 minutes for trips from EBD stations and 31.8 minutes for those from control stations. The average fare is 5.00 RMB in the EBD sample and 5.14 RMB in the control sample, reflecting both the distance-based pricing schedule and the EBD discount. The signed rescheduling

Table 2: Summary Statistics

	EBD Stations		Control Stations	
Variable	Mean	SD	Mean	SD
# OD pairs	1,621		1,920	
# markets	17,777		22,414	
# passengers	10.96	21.81	12.41	16.28
# frequent passengers	10.23	20.96	11.57	15.62
Distance (km)	20.03	7.97	18.84	6.93
Total travel time (min)	46.93	18.21	41.57	14.49
In-train travel time (min)	31.95	13.82	31.75	12.92
Price (yuan)	5.00	0.96	5.14	0.78
Rescheduling (min, raw)	−2.19	8.88	−0.15	2.32
Crowding K (person·min/m^2)	121.74	59.36	123.35	45.25
Crowding density k (person/m^2)	3.88	1.35	4.14	1.34
Crowding density from ACP (infrequent)	0.07	0.05	0.07	0.07
Crowding density from ACP (total)	0.70	0.59	0.84	0.85
Crowding experience from ACT (infrequent)	1.23	0.95	1.09	1.02
Crowding experience from ACT (total)	11.55	9.85	12.42	12.16

Notes: Each observation is a product (jdt), defined as an OD pair–date–15-minute time bin. There are 142,216 observations in EBD covered OD pairs and 179,312 in control OD pairs. All variables below the second horizontal rule are weighted by ridership. ACP denotes accidental companion passengers; crowding density from ACP has a unit of persons/m^2. ACT stands for accidental companion time; crowding experience from ACT has a unit of person·min/m^2. "Infrequent" restricts to passengers with no more than three trips during the week.

deviation averages −2.19 minutes in the EBD sample (SD 8.88) and −0.15 minutes in the control sample (SD 2.32).[10] The negative EBD mean is consistent with riders advancing their departures to qualify for the discount.

Passengers in the sample experience substantial crowding. The ridership-weighted average in-train crowding density is 3.88 persons per square meter for trips originating from EBD stations and 4.14 for those from control stations. The total in-train crowding experience, which integrates real-time crowding density over the duration of the trip, averages 121.7 person-minutes per square meter for EBD stations and 123.4 for control stations.[11]

Accidental companion passengers contribute meaningfully to crowding. For trips from EBD stations, total ACPs contribute a crowding density of 0.70 persons per square meter, accounting for

[10]The signed deviation is near-zero in the control sample by construction; the small residuals reflect the use of predicted travel time $\widehat{\mathrm{TT}}_{jdt}$ in place of realized values, so the identifying variation in ρ comes predominantly from the EBD sample.

[11]The product of the average density and the average in-train time does not match the average integrated crowding exactly, because the integrated crowding is a sum of trip-level products of density and time, while the table reports trip-level means of each component separately; equivalently, $\overline{k \cdot t} \neq \bar{k} \cdot \bar{t}$ unless k and t are uncorrelated across trips.

roughly 18% of measured crowding. Infrequent riders, our preferred set of ACPs, contribute 0.07 persons per square meter. When weighted by the duration of shared travel, the accidental companion time of infrequent riders averages 1.23 person-minutes per square meter. The control sample has comparable figures.

4 Stylized Facts

This section documents several empirical patterns that motivate our modeling and identification choices. We show that the EBD policy effectively creates bunching in departure times and affects crowding, that crowding is predictable from regular commuting patterns, but it does not cause meaningful delays or rerouting. These facts support treating crowding as a pure disamenity and validate our crowding measures, which assumes that passengers take the fastest route.

4.1 The EBD Creates Bunching in Departure Time and Crowding

Figure 1 plots the number of passengers entering, exiting, transferring through, and riding in trains at EBD and control stations by 5-min time bin during morning hours. At EBD stations (panel a), the entry flow displays a noticeable discontinuity around 7 AM. A cluster of passengers tap in just before the cutoff to qualify for the 30% discount, creating a visible dent in entries immediately afterward. This bunching in departure times in turn disturbs the distribution of passengers passing through the station around 7 AM. Because passengers passing through the station boarded the train at upstream stations and are themselves responding to the EBD cutoff, the resulting dent in the in-train crowding density (Figure 2) is less sharp than the discontinuity in entry flows. The diffusion across upstream stations smooths the effect, but it also means that the EBD generates changes in crowding that ripple beyond the immediate neighborhood of the cutoff time. A passenger scheduled for an 8 AM departure who has no interest in taking the early-bird discount may still adjust her departure time because the crowding on her regular trip is altered by other passengers responding to the EBD.

In contrast, panel (b) of Figure 1 shows that control stations exhibit no such discontinuity. The entry flow rises smoothly through the morning peak with no bunching around 7 AM, confirming that the EBD is the source of the departure-time distortion observed at treated stations.

4.2 Crowding Is Predictable

A key feature of the demand model is that passengers choose departure times based on *expected* crowding, which we proxy with fitted values from regressions of realized crowding on OD-by-time-bin fixed effects. This approach requires that crowding and ridership follow sufficiently regular patterns for experienced commuters to anticipate. Table 3 confirms this premise. The upper panel

reports the R^2 from regressions of realized crowding on progressively finer sets of fixed effects. With OD fixed effects alone, the R^2 for crowding experience (K, measured in persons·mins/m^2) is 0.83. Adding the time-bin dimension raises the R^2 to 0.99, and further conditioning on day of week yields an R^2 above 0.99. The crowding density (k, measured in persons/m^2) is similarly predictable, with an R^2 of 0.98 under OD-time fixed effects. These high R^2 values indicate that nearly all variation in crowding is explained by the route and time of day, regular commuters are likely aware of the crowding situation.

The lower panel reports analogous regressions for ridership. Among frequent riders, who average at least three subway trips per week, R^2 values reach 0.93 under OD-time fixed effects and 0.94 under OD-day-of-week-time fixed effects. Infrequent riders are less predictable, with an R^2 of 0.52 under OD-time effects and 0.63 under OD-day-of-week-time effects, consistent with their more irregular travel patterns. The high predictability of ridership among frequent riders, combined with the near-perfect predictability of crowding, supports the assumption that passengers form accurate expectations about the crowding they will encounter on a given route and time.

Table 3: Predictability of Crowding and Ridership

	Crowding Experience (K)			Crowding Density (k)		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Crowding						
R^2	0.8301	0.9876	0.9916	0.6511	0.9812	0.9870
Obs. (million)	5.79	5.74	4.96	5.79	5.74	4.96
	Frequent Riders			Infrequent Riders		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: Ridership						
R^2	0.6804	0.9342	0.9449	0.4317	0.5226	0.6281
Obs. (million)	12.53	12.53	11.93	12.53	12.53	11.93
OD FE	X			X		
OD-Time FE		X			X	
OD-DOW-Time FE			X			X

Notes: Panel A reports R^2 from regressions of realized crowding on fixed effects of increasing granularity. Crowding experience K is measured in person·min/m^2 and crowding density k in persons/m^2. Panel B reports analogous regressions for ridership, separately for frequent riders (≥4 trips per week) and infrequent riders. High R^2 values confirm that crowding and ridership are largely predictable from regular commuting patterns.

4.3 Crowding Does Not Slow Down the Train

Excessive time cost is a potential consequence of crowding and a confounding factor in the WTP to avoid crowding. In many transit settings, such as bus networks, high crowding coincides with road congestion, making it difficult to separate disutility from crowding per se from disutility caused by longer travel times. A distinctive feature of the current empirical setting is that in-train crowding does not meaningfully affect train speed, as trains run on dedicated tracks with fixed schedules. This makes the Beijing Subway an ideal setting for identifying the pure disamenity cost of crowding. We test this premise directly.

Table 4 reports quantile regressions of individual trip travel time on in-train crowding density and in-station ridership, controlling for OD-day and line-time fixed effects. Each observation is an individual trip. All columns have 33,436,600 observations. The results show that increasing crowding density by one person per square meter increases travel time by only 0.23 to 0.46 minutes, depending on the quantile. Given that average travel times range from 35 to 57 minutes across these quantiles, these effects represent less than a 1% increase in total travel time.

All specifications control for in-station ridership (persons). This control ensures that the coefficients on crowding density capture the relationship between *in-train* crowding and travel time, which is the margin relevant for the demand model, rather than confounding it with delays caused by congestion in station corridors, platforms, and gates.

Table 4: The Effects of Crowding on Travel Time

	Travel Time					
	p10	p25	p50	p75	p90	p99
	(1)	(2)	(3)	(4)	(5)	(6)
Crowding density k (persons/m^2)	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 (persons)	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	35.10	36.90	38.59	40.55	43.52	56.88
Mean crowding density k (persons/m^2)	2.87	2.87	2.87	2.87	2.87	2.87
Mean in-station ridership (persons)	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: Quantile regressions of trip travel time on in-train crowding density (persons/m^2) and in-station ridership (persons). Each observation is an individual trip. All columns have 33,436,600 observations. Standard errors in parentheses, clustered at the OD level.

To put the magnitude in perspective, we convert the delay cost into monetary terms. The median monthly earnings for full-time workers is approximately 8,000 RMB. Assuming 160 work hours per month, this implies an earning rate of about 0.8 RMB per minute. Assuming the value of time during a subway trip is 50% of the wage rate, the value of time in the subway is 0.4 RMB per minute. At

the 90th percentile of travel time, increasing crowding density by one person per square meter slows down an average trip by 0.33 minutes, which translates to a monetary cost of less than 0.15 RMB. As we show in the estimation results, this is an order of magnitude smaller than the willingness to pay for reducing the discomfort from crowding.

The relationship between crowding and delay is slightly convex. Table 5 reports results from a quadratic specification. The linear term is negative (−0.16) and the quadratic term is positive (0.05), indicating that crowding has essentially no effect on travel time at low density levels and a small positive effect at high density levels. Figure 7 plots the implied total monetary cost of crowding-induced delay as a function of crowding density, evaluated at the median travel time. Below the 90th percentile of crowding density, the delay cost remains below 0.3 RMB, a tiny fraction of the average fare and the WTP for crowding reduction estimated later.

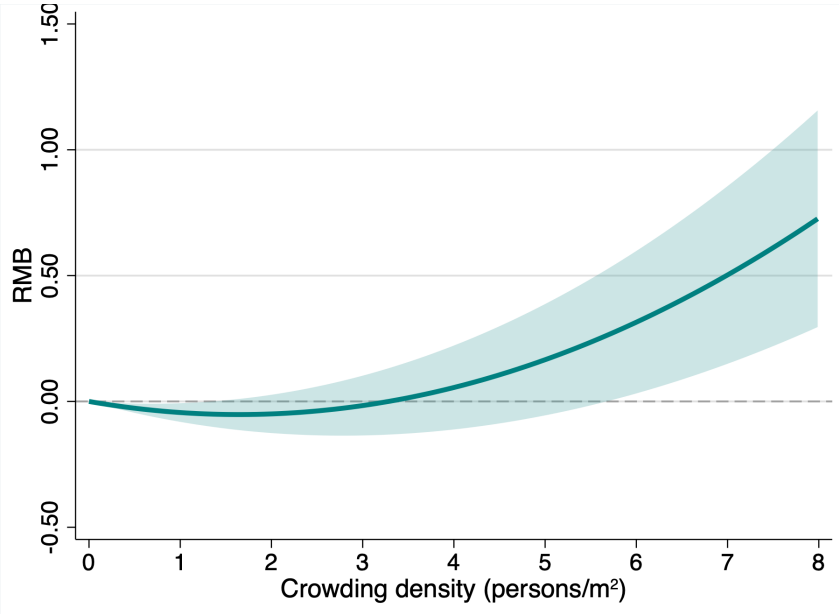
Table 5: Crowding and Travel Time: Quadratic Specification

	(1) Travel Time
Crowding density k (persons/m^2)	-0.1586
	(0.0468)
Crowding density k squared	0.0483
	(0.0063)
In-station ridership (persons)	0.0037
	(0.0001)
N	30,280,427
Mean dep. var.	38.05
Mean crowding density k (persons/m^2)	2.87
Mean in-station ridership (persons)	667.72
OD-Day FE	X
Line-Time FE	X

Notes: The quadratic specification includes crowding density and its square. 30,280,427 observations (cells with fewer than 2 trips are excluded). Standard errors clustered at OD level.

In-station crowding is also positively associated with total travel time, though it is not the focus of this paper. A one-unit increase in in-station ridership (one additional person) is associated with a 0.003 to 0.005 minute increase in travel time, depending on the quantile. To compare the magnitudes, note that an increase of one person per square meter in in-train crowding density represents about a 25% increase from the ridership-weighted sample average of 4.0 persons/m^2. A comparable 25% increase from the mean in-station ridership of 668 persons corresponds to an increase of 167 persons. At the 90th percentile, this translates to an increase in total travel time of $167 \times 0.004 = 0.67$ minutes, or approximately 0.3 RMB in value of time. The fact that the same proportional increase in in-station crowding produces a larger delay than the same proportional increase in in-train crowding suggests that crowding delays travel primarily through slower walking traffic inside stations rather than by slowing trains down. This is intuitive, as pedestrian flows through corridors, stairways, and platform

Figure 7: Total Monetary Cost of Crowding-Induced Delay



Notes: This figure plots the implied delay cost (in RMB) from the quadratic specification as a function of crowding density. Value of time is assumed to be 50% of median earnings (0.4 RMB/min). Delay cost remains below 0.3 RMB for crowding levels up to the 90th percentile.

gates are directly impeded by larger crowds, whereas trains operate on fixed schedules regardless of passenger load.

While this paper focuses on in-train crowding, we note that in-station crowding also causes discomfort and generates disutility. In-station and in-train crowding are positively correlated, as a passenger experiencing a crowded platform can generally expect a crowded train. Although we do not directly estimate the disutility from in-station crowding, we show in the robustness checks that in-station crowding does not confound the estimated WTP to avoid in-train crowding.

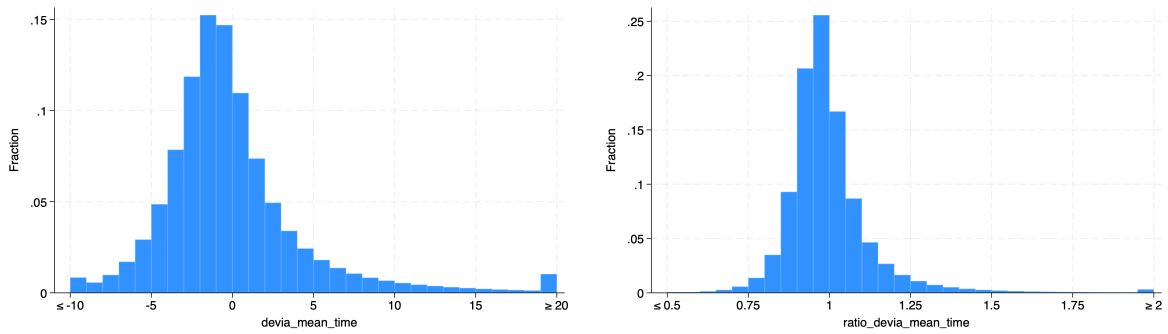
4.4 Crowding Does Not Result in Rerouting

The computation of in-train crowding relies on a train assignment algorithm that assigns each passenger to the fastest route. A natural concern is that some passengers may deviate from this route in response to crowding. If passengers systematically choose longer but less crowded alternatives, the measured crowding on their assigned route would overstate the crowding they actually experience, and the degree of overstatement would increase with crowding levels. We present two related pieces of evidence that crowding does not cause meaningful rerouting.

First, Figure 8 shows the distribution of trip-level deviations from the OD-specific mean travel time. Panel (a) reports the deviation in minutes; panel (b) reports the ratio of individual travel time to the OD mean. The distributions are tightly concentrated around zero. The vast majority of trips

cluster close to the mean, with 65.1% of trips falling within 3 minutes, 83.6% within 5 minutes, and 95.8% within 10 minutes. When expressed as a ratio to the mean, 71.6% of trips fall in the interval [0.9, 1.1] and 94.8% in [0.75, 1.25]. If a substantial fraction of riders were taking alternative routes, we would expect to see a cluster of travelers in the far right tail, as rerouting causes considerable delay. No such cluster is apparent.

Figure 8: Histograms of Deviation from the Mean Travel Time



(a) Absolute Deviation (Minutes)

(b) Ratio to Mean Travel Time

Notes: Panel (a) shows the absolute deviation of each trip's travel time from the OD-specific mean (in minutes). Panel (b) shows the ratio of trip travel time to the OD-specific mean. Sample includes all trips in the working sample.

Rerouting takes considerably longer than the shortest path, so if crowding systematically causes rerouting, it should have a large effect on the dispersion of travel time. Appendix Figure B.4 shows the distribution of travel time as a function of crowding. As in-train crowding increases, the distribution of trip time becomes slightly wider, indicating a larger variance, but the increase is not visually salient and there is no cluster of riders in the right tail at higher crowding quartiles. Much of the widening is driven by crowded stations slowing down foot traffic rather than by larger crowds slowing down trains. For trips that do not require a transfer, the distribution is much tighter and crowding does not lead to visually discernible increases in the variance. Trips requiring one or more transfers have larger variances in travel time to begin with, and crowding produces somewhat more visible increases in spread. Crucially, even for trips with transfers, the distributions remain similar across crowding quartiles, indicating that the right tail grows with the *number of transfers* rather than with *crowding levels*. Since trips without transfers have only one feasible route, any deviation from the mean must reflect train-waiting time variation or measurement error rather than rerouting.

Second, we show formally in regressions that crowding has a negligible impact on the travel time variance. Table 6 tests this directly. The dependent variables are the squared deviation (column 1), absolute deviation (column 2), and ratio to mean (column 3) of each trip's travel time from its OD-date-time-bin cell mean. All regressions control for OD-day and line-time fixed effects and include both in-train crowding density and in-station ridership.

The coefficients on crowding density are statistically significant but economically small. A

Table 6: The Effects of Crowding on Travel Time Variance

	(1) Squared Deviation	(2) Absolute Deviation	(3) Ratio to Mean
Crowding density k (persons/m^2)	0.6014	0.1244	-0.0000
	(0.1289)	(0.0090)	(0.0000)
In-station ridership (persons)	0.0023	0.0004	0.0000
	(0.0003)	(0.0000)	(0.0000)
N	30,280,427	30,280,427	30,280,427
Mean dep. var.	16.38	2.40	1.00
Mean crowding density k (persons/m^2)	2.87	2.87	2.87
Mean in-station ridership (persons)	667.72	667.72	667.72
OD-Day FE	X	X	X
Line-Time FE	X	X	X

Notes: Each observation is a trip. The dependent variable in column (1) is the squared deviation of the trip's travel time from its cell mean, $(TT_i - \bar{TT}_{\text{cell}})^2$, where a cell is defined as an OD-date-15-minute-bin. Column (2) uses the absolute deviation $|TT_i - \bar{TT}_{\text{cell}}|$. Column (3) uses the ratio of the trip's travel time to its cell mean, $TT_i/\bar{TT}_{\text{cell}}$. Cells with fewer than 2 trips are excluded, yielding 30,280,427 observations. Standard errors in parentheses, clustered at the OD level.

one-unit increase in crowding density (one person/m^2) is associated with a 0.12-minute increase in absolute deviation (column 2), relative to a mean absolute deviation of 2.40 minutes. At the ridership-weighted sample mean crowding level of 4.0 persons/m^2, crowding contributes approximately $0.12 \times 4.0 \approx 0.48$ minutes to the absolute deviation, or about 20% of the baseline. The within-R^2 of these regressions is essentially zero, indicating that crowding explains a negligible share of the trip-level variation in travel times. In column 3, the correlation between crowding and ratio to the mean is essentially zero. These patterns are inconsistent with systematic rerouting, which would generate large right-tail outliers and substantial increases in travel time dispersion.

5 Estimation Results

5.1 GMM Estimates with Preferred IV

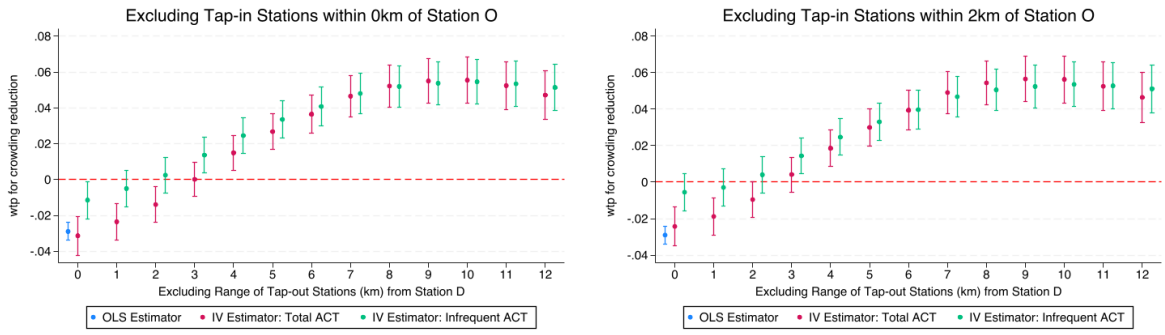
Table 7 reports the estimation results of Equation (4) for the sample of frequent riders. Column (1) presents OLS estimates with OD-day and time-day fixed effects. The fare coefficient is negative and precisely estimated, indicating that passengers are less likely to choose a time bin with a higher fare. The rescheduling coefficient is also negative. The crowding coefficient, however, is *positive*, and the implied MWTP for crowding reduction is correspondingly negative. Rather than reflecting a genuine preference for crowded trains, the wrong sign highlights the endogeneity of crowding: trips taken at crowded times and on crowded routes carry high unobserved value, for instance because passengers sharing the same route need to arrive at work at the same hour. The positive correlation between crowding and unobserved demand swamps the true disutility, biasing the OLS estimate of

β upward.

To address the endogeneity of crowding, we exploit plausibly exogenous variation generated by accidental companion passengers (ACPs), as described in Section 3.3. The potential endogeneity arises from demand shocks correlated at the origin, the destination, or through common travel purpose. A good candidate for an instrument variable can use trips with different origins and destinations by different types of users, while still relevant for the experienced crowdedness by the focal passenger. But how different should valid ACPs be? To disentangle these sources, we vary the construction of the instrumental variable by progressively tightening restrictions on which co-riders qualify as ACPs.

Figure 9 presents the estimated MWTP for crowding reduction across specifications that vary the minimum distance $|D - D'|$ between the focal passenger's destination D and the ACP's destination D'. Each panel plots estimates from three models: OLS, GMM with ACT constructed from all ACPs, and GMM with ACT restricted to infrequent ACPs.

Figure 9: Source of Crowding Endogeneity



(a) No restriction on ACP origin

(b) Excluding ACP origins within 2 km of O

Notes: Each point reports the estimated MWTP for reducing crowding by one person per square meter for one minute (RMB), from a separate regression. The horizontal axis varies the minimum distance $|D - D'|$ between the focal passenger's destination D and each ACP's destination D'. Blue diamonds are OLS estimates, which shows a precisely estimated *negative* MWTP of about 3 cents RMB to reduce crowding density by one person per square meter for one minute; red circles use ACT constructed from all ACPs; green triangles restrict ACT to infrequent ACPs ($\leq$ 3 trips/week). Vertical bars show 95% confidence intervals. Standard errors clustered at the OD level.

Panel (a) requires ACPs to board an upstream station than the focal passenger, but imposes no other restrictions on the origin station for the ACPs. When we only require $D \neq D'$ but distance $|D - D'|$ is not restricted (the leftmost point), the GMM estimate with ACT as IV yields an MWTP to be negative, with a magnitude similar to the OLS. Using only infrequent ACPs pushes the estimate upward, but the MWTP remains statistically negative and statistically significant. As $|D - D'|$ increases, gradually removing ACPs whose destinations are close to D, the GMM estimate rises steadily. By restricting $|D - D'|$ to exceed 3 km, the MWTP for crowding reduction becomes positive. Restricting $|D - D'|$ to be greater than approximately 7 to 8 km, the estimates based on all ACPs and

those restricted to infrequent ACPs converge. The fact that estimates using all ACPs and infrequent ACPs as the IV converge as $|D - D'|$ increases is intuitive. Even ACPs are traveling for the same purpose (say arriving at work on time), the fact that they are going to a faraway destination means that they are responding to an unrelated demand shock. When $|D - D'|$ is greater than 8 km, the MWTP stabilizes at roughly 0.05 RMB per person·min/m^2. Further tightening the distance restriction does not meaningfully change the estimates.

Panel (b) repeats the exercise after additionally excluding ACPs that originate from a station O' within 2 km of the focal passenger's origin O. The results are nearly identical to those in panel (a), suggesting that the source of endogeneity is not from correlated demand shocks at the origin. Taken together, the two panels show that the endogeneity of crowding stems primarily from common shocks at the *destination*. Intuitively, the most important shock that drives crowded subway cars is the need to arrive at work at a required time, which concentrates travel demand among passengers heading to nearby destinations during the same narrow window.

Based on the convergence patterns in Figure 9, our preferred instrument restricts ACPs to passengers (O', D') who satisfy three criteria: (a) those who board at an upstream station from O; (b) $\text{dist}(D', D) \geq 8$ km, ensuring the ACP's destination is sufficiently far from D; and (c) infrequent users who take no more than three subway trips in the week. The overlap between each qualifying ACP and the focal trip yields the accidental companion time (ACT), which serves as the instrumental variable for expected crowding.

Requiring $|D - D'| \geq 8$ km carves out an exclusion zone around D as big as 200 square kilometers. One may worry that it leaves too few qualifying ACPs to generate a strong first stage. This concern is mitigated by the scale of the Beijing Subway network. Figure B.5 in Appendix B shows that the EBD stations are at the far suburban ends of the network, while the vast majority of the system extends well beyond the 8-km radius from the city center (panel a). Because destinations are spread throughout this extensive network, the 8-km exclusion zone around each focal destination D removes only a small share of potential ACPs. Indeed, the median infrequent ACP alights at a station roughly 11 km from D (panel b), well beyond the exclusion threshold. As a result, the first-stage relationship remains strong, with Wald F-statistics exceeding 280 in all specifications (Table 7).

In addition to ACT, we include four instruments based on the EBD policy to form the multi-IV specification. Additional instruments are needed to estimate the random coefficient model. These are interactions between the treatment indicator (an EBD station) and a pre-7 AM indicator as well as linear, quadratic, and cubic time trends:

$$\text{Treat}_j \times 1\{t < 7{:}00\}, \quad \text{Treat}_j \times f(t), \quad \text{Treat}_j \times f(t)^2, \quad \text{Treat}_j \times f(t)^3.$$

These instruments exploit the temporal reallocation of trips induced by the fare discount, generating exogenous variation in crowding that is complementary to the network-based ACT instrument.

The remaining columns of Table 7 report GMM estimates in which crowding is instrumented by ACT. Columns (2) and (3) use a single IV (ACT alone) and report the first-stage and second-stage results, respectively. Columns (4) and (5) use the multiple IV specification (ACT plus four EBD policy interactions) and again report the first and second stages. Column (6) extends the multi-IV specification by allowing for a non-linear crowding effect through a quadratic term. All columns include OD-day and time-day fixed effects.

Table 7: OLS and GMM Estimation Results

	OLS	GMM: Single IV		GMM: Multiple IV		GMM: Nonlinear
		1st Stage	2nd Stage	1st Stage	2nd Stage	2nd Stage
	(1)	(2)	(3)	(4)	(5)	(6)
Crowding	0.0062		−0.0201		−0.0094	
	(0.0003)		(0.0019)		(0.0011)	
Crowding^2						−0.000028
						(0.000003)
Price	−0.2093	−5.1519	−0.3280		−0.1821	−0.1513
	(0.0151)	(0.4331)	(0.0196)		(0.0094)	(0.0090)
Rescheduling	−0.0299	0.0354	−0.0274	−0.1360	−0.0250	−0.0252
	(0.0010)	(0.0306)	(0.0014)	(0.0243)	(0.0010)	(0.0010)
ACT		11.5338		11.1407		
		(0.4788)		(0.4645)		
Other IVs				✓		
Wald F-Test			580.2		284.8	441.8
R^2	0.6588		−0.3272		−0.0720	
OD-Day FE	X	X	X	X	X	X
Time-Day FE	X	X	X	X	X	X
Mean Rescheduling (mins)	4.654		4.654		4.654	
Mean Crowding (person·min/m^2)	127.1		127.1		127.1	
Implied MWTP						
Crowding	−0.0296		0.0613		0.0516	*see*
	(0.00266)		(0.00550)		(0.00613)	*Fig. B.6*
Rescheduling	0.143		0.0835		0.137	0.1662
	(0.0114)		(0.00689)		(0.00894)	(0.0120)

Notes: This table reports OLS and GMM estimates of Equation (4) for the sample of frequent riders. All specifications include OD-day and time-day fixed effects. Column (1) is estimated by OLS. In columns (2)–(3), crowding is instrumented by ACT from infrequent riders with $|D - D'| \geq 8$ km. In columns (4)–(6), four EBD policy interactions are added as instruments (Treat $\times 1\{t < 7{:}00\}$, Treat $\times$ trend, Treat $\times$ trend^2, Treat $\times$ trend^3). Column (6) includes a quadratic crowding term. Standard errors in parentheses, clustered at the OD level. MWTP computed as the ratio of the relevant coefficient to the price coefficient.

In the first stage (columns 2 and 4), ACT is a strong positive predictor of crowding, with coefficients of 11.5 and 11.1 and first-stage Wald F-statistics of 580 and 285 in the single-IV and multi-IV specifications, respectively. The F-statistics comfortably exceed conventional thresholds for weak instruments.

Turning to the second stage, the crowding coefficient $\hat{\beta}$ is *negative* in all GMM specifications, in sharp contrast to the positive OLS estimate in column (1). With a single IV (column 3), the crowding

coefficient is -0.0201 and the price coefficient is -0.0328, yielding an implied MWTP for crowding reduction of 0.061 RMB per person·min/m^2. The implied MWTP to reduce rescheduling by one minute is 0.0835 RMB. The multi-IV specification (column 5) produces a similar implied MWTP of 0.052 RMB for crowding reduction and 0.137 RMB for rescheduling reduction. The multiple-IV specification is our preferred specification.

To gauge the economic magnitudes, consider a representative trip lasting approximately 32 minutes in train. The willingness to pay to reduce crowding density by one person per square meter over the course of the trip is $0.052 \times 32 \approx 1.7$ RMB, or roughly 35% of the average fare of 5 RMB. The total crowding disutility for a trip at the mean total crowding of 122 person·min/m^2 amounts to $0.052 \times 122 \approx 6.3$ RMB, exceeding the fare itself. These magnitudes confirm that in-train crowding is a first-order determinant of the value of a subway trip. For rescheduling, deviating from the optimal arrival time by 15 minutes costs $0.137 \times 15 \approx 2.1$ RMB. Relative to the median hourly wage of about 44 RMB, the rescheduling cost of a 15-minute deviation amounts to about 5% of hourly earnings. The relatively small rescheduling cost may be due to the fact that most rescheduling in our setting is departing and arriving earlier than optimal. Rescheduling to a later arrival time may incur a larger cost. Nevertheless, those findings suggest that passengers face a meaningful tradeoff between departing at a less crowded time and arriving closer to their preferred schedule.

Column (6) allows for a non-linear effect of crowding by including the quadratic term of the crowding experience. The estimated coefficient on the quadratic term is negative and precisely estimated, indicating that the marginal disutility of crowding is increasing and convex. The MWTP for crowding reduction varies with the level of crowding, as shown in Figure B.6 in Appendix B. At the 25th percentile of crowding density, the MWTP is 0.032 RMB per person·min/m^2, rising to 0.048 at the median and 0.086 at the 95th percentile (around 6 persons/m^2). The increasing pattern implies that the welfare cost of crowding is concentrated on the most congested trips.

5.2 Robustness

ACP as Instrument. Our preferred instrument uses accidental companion *time* (ACT), which weights each co-rider by the duration of shared travel. A natural alternative is to use the raw count of accidental companion passengers (ACP), discarding the time weights. Table B.2 in Appendix B reports both single-IV and multi-IV specifications using ACP. In the first stage, ACP is a strong predictor of crowding, with a coefficient of approximately 180 and Wald F-statistics of 461.9 (single-IV) and 288.3 (multi-IV). The MWTP for crowding reduction is 0.058 RMB per person·min/m^2 under single-IV ACP and 0.047 under multi-IV ACP. These estimates are close to their ACT counterparts (0.061 and 0.052, respectively), as expected because ACP and ACT exploit the same source of exogenous variation. The stability of the crowding WTP across the two instrument constructions is reassuring.

Controlling for In-Station Crowding. A potential concern is that in-train crowding may be correlated with in-station congestion, which could independently affect route choice. In-station crowding is qualitatively different from in-train crowding: passengers experience station congestion in open corridors, platforms, and gate areas where the per-capita space is substantially larger than inside a train car. The main channel through which station congestion could plausibly affect riders is via delays in gates, corridors, and platforms; Section 4.3 shows directly that any such delay is small in magnitude in our setting (an additional person per square meter of in-station ridership raises total travel time by less than one minute even at the 90th quantile). To examine the disamenity channel itself, Table B.3 in Appendix B reports four specifications. We use a unified definition of in-station crowding throughout: InStn_{jdt} is the total in-station ridership-time per product, summed across all riders in the product cell of their individual entry, transfer, and exit times, in thousands of person-minutes.

Column (1) is the single-IV ACT baseline. Column (2) adds InStn as an uninstrumented control. The in-station coefficient is positive (+0.040), and the in-train MWTP shifts from 0.061 to 0.050. A positive coefficient on a disamenity indicates unaddressed endogeneity: high-demand markets attract both more in-station congestion and more riders, biasing an uninstrumented coefficient on InStn upward. Column (3) is the preferred multi-IV baseline. Column (4) treats InStn as a second endogenous regressor, jointly instrumented with in-train crowding by the multi-IV set (ACT + four EBD policy interactions). The EBD policy instruments shift entry timing, which moves both in-train crowding and in-station ridership through the same exogenous mechanism, providing the variation needed to identify both endogenous regressors. The Wald F-statistic is 470.6, above conventional thresholds, and the in-train crowding coefficient is -0.0078, implying an MWTP of 0.048 RMB per person·min/m^2 (within roughly 7% of the multi-IV baseline). The in-station coefficient turns negative (-0.008 per thousand person·min), reversing the positive sign in column (2) and confirming that the column-(2) coefficient reflected endogeneity rather than a true positive amenity. The stability of the in-train MWTP between the multi-IV baseline and column (4) confirms that omitting in-station crowding does not substantially bias our main estimates.

The in-station coefficient also yields an interpretable secondary magnitude. The implied WTP for in-station crowding reduction is $0.008/0.162 \approx 0.048$ RMB per thousand person·min reduction. Evaluated at the sample mean (InStn ≈ 14.3 thousand person·min), the implied disutility from average in-station crowding is roughly $0.048 \times 14.3 \approx 0.69$ RMB, an order of magnitude smaller than the average in-train disutility computed in Section 5.1, and one of the reasons we focus the welfare analysis on the in-train margin.

5.3 Heterogeneous Willingness to Pay

The standard logit model restricts all passengers to share the same preference parameters. We now relax this assumption and allow the marginal utility of price, crowding, and rescheduling to vary with income, as specified in the random coefficient model of Equation (5). Each preference coefficient is a linear function of simulated passenger income, recentered at the sample mean of 8.77 thousand RMB. The model is estimated by GMM with the same multi-IV instrument set used in the preferred specification, following the BLP contraction mapping procedure described in Section 2.1.[12]

Table 8 reports the estimation results. The price intercept $\hat{\alpha}_0$ = -0.179 indicates that the average passenger is price-sensitive, while the positive income interaction $\hat{\alpha}_1$ = 0.013 implies that higher-income riders are less responsive to fare changes. The crowding intercept $\hat{\beta}_0$ = -0.026 confirms that crowding reduces utility, and the negative income interaction $\hat{\beta}_1$ = -0.009 reveals that higher-income passengers experience substantially stronger crowding disutility. The rescheduling intercept $\hat{\rho}_0$ = -0.017 is negative, but the income interaction $\hat{\rho}_1$ = 0.002 is not statistically different from zero, suggesting that the cost of deviating from the optimal arrival time is relatively uniform across income levels.

The income gradient in both the price coefficient and the crowding coefficient produces sharply heterogeneous willingness to pay for crowding reduction. Because the MWTP is the ratio β_i/α_i, two forces work in the same direction for higher-income passengers: they dislike crowding more (a more negative β_i) and are less price-sensitive (a less negative α_i). Figure 10 plots the estimated MWTP for crowding reduction and rescheduling avoidance as a function of monthly income, with 95 percent confidence intervals computed by the delta method.

The MWTP for crowding reduction is steeply increasing in income (panel a). At the median income of 7.5 thousand RMB, the MWTP is approximately 0.07 RMB per person·min/m^2, close to the homogeneous estimate of 0.052 from Table 7. At the 90th percentile of income (17.5 thousand RMB), the MWTP rises to roughly 1.65 RMB per person·min/m^2, more than twenty times the median value. In contrast, the MWTP for rescheduling avoidance is relatively flat across income levels (panel b), hovering around 0.10 RMB per minute throughout most of the income distribution. The stability of the rescheduling WTP is consistent with the insignificant income interaction $\hat{\rho}_1$ and suggests that the cost of arriving at an inconvenient time is not strongly differentiated by income.

The steep income gradient in crowding WTP has important implications for policy. It suggests that higher-income passengers would be willing to pay a substantial premium for less crowded ser-

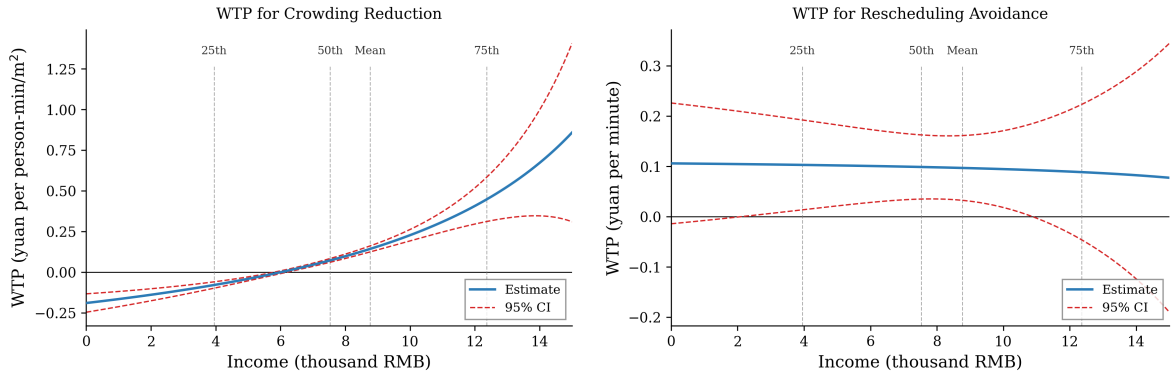
[12]The random-coefficient model is estimated on the *overall* sample of riders with OD-day and time-bin fixed effects, rather than on the frequent-rider sub-sample with time-by-day fixed effects used as the preferred constant-coefficient specification. The mean coefficients are stable to within roughly 10% across these alternatives in the constant-coefficient model (Table 7), so the BLP intercepts $\alpha_0, \beta_0, \rho_0$ would not shift materially under the more saturated specification. The choice to estimate on the overall sample is motivated by the outside option, which is defined for the full commuting population via the BHTS-imputed mode shares; a frequent-rider restriction would require an ad hoc "frequent-commuter market size" for which we have no clean empirical anchor.

Table 8: Random Coefficient Model Estimation Results

	(1)
Price (α_0)	−0.1787
	(0.0109)
Price × Income (α_1)	+0.0131
	(0.0048)
Crowding (β_0)	−0.0256
	(0.0006)
Crowding × Income (β_1)	−0.0093
	(0.0002)
Rescheduling (ρ_0)	−0.0173
	(0.0058)
Rescheduling × Income (ρ_1)	+0.0016
	(0.0019)
GMM Objective	0.2599
OD-Day FE	X
Time FE	X

Notes: This table reports the random coefficient model estimates of Equation (5). Income is measured in thousands of RMB per month and recentered at the sample mean (8.77 thousand RMB). All specifications include OD-day and time fixed effects. Standard errors in parentheses. The model is estimated using the multi-IV instrument set (ACT from infrequent riders with $|D - D'| \geq 8$ km, plus four EBD policy interactions).

Figure 10: Willingness to Pay by Income



(a) WTP for Crowding Reduction

(b) WTP for Rescheduling Avoidance

Notes: The solid line shows the point estimate of MWTP as a function of monthly income. Dashed lines indicate the 95% confidence interval (delta method). Vertical dashed lines mark selected percentiles of the income distribution. Panel (a) reports the MWTP for reducing in-train crowding by one person·min/m^2 (in RMB). Panel (b) reports the MWTP for reducing schedule deviation by one minute (in RMB). The plots are truncated at 15 thousand RMB (between the 75th and 90th percentiles), beyond which the price coefficient $\alpha(I)$ approaches zero and the WTP ratio becomes imprecisely estimated.

vice, while lower-income passengers are relatively indifferent to crowding. This pattern provides a potential rationale for service differentiation, where higher-income riders self-select into a more expensive but less crowded option. We explore this possibility in the welfare analysis of Section 6.

6 Welfare and Counterfactual

The structural parameters from Section 5 allow us to quantify the crowding externality and evaluate alternative policies designed to internalize it. Subway crowding is a congestion externality in the classical sense. Each additional rider imposes a disamenity on fellow passengers that is not incorporated in the fare, so the competitive equilibrium features inefficient over-crowding. In this section, we use the estimated demand system to compute the socially optimal level of ridership and compare three policy instruments that can close the gap between private and social demand. A Pigouvian crowding tax prices the externality directly. Capacity-rationing queuing targets the optimal ridership through a non-price mechanism. Finally, a two-class train configuration with differentiated pricing exploits the steep income gradient in crowding valuations documented in Section 5.3.

6.1 Setup

We study a single origin-destination segment of the subway, on which every passenger imposes crowding on every other passenger for the duration of the trip. There are two income groups $g \in \{\ell, h\}$ with equal population $M_\ell = M_h = M/2$, where M denotes the total market size. A passenger either rides the subway or takes the outside option, the utility from which is standardized to one. Deterministic utility from riding is

$$V_g \;=\; A \;+\; \alpha_g P \;+\; \beta_g \underbrace{\frac{\text{TravelTime}}{\text{CarArea} \times J}}_{L} N, \tag{15}$$

where A is a common mean utility, P is the fare, N is total ridership on the segment. Let $L = \text{TravelTime}/(\text{CarArea} \times J)$, where J is the number of cars in a train, thus the denominator indicates the total floor space in a train. NL measures the total crowding exposure a passenger experiences during the ride (crowding density times minutes). Passengers derive disutility from price and crowding, $\alpha_g < 0$ and $\beta_g < 0$. Idiosyncratic taste shocks are i.i.d. type-I extreme value, so the logit share for group g is $s_g(V_g) = e^{V_g}/(1 + e^{V_g})$ and group ridership is $N_g = M_g\, s_g(V_g)$, with $N = \sum_g N_g$.

We split riders at the 41st and 59th percentiles of the simulated income distribution from the Baidu commuting data.[13] The low-income group has a monthly income of y_ℓ=6.19 thousand RMB,

[13]The split p41/p59 is symmetric around the median (41 + 59 = 100). The 41st percentile is chosen because it sits just above the point at which the crowding coefficient changes sign, so that low-income riders have a weakly non-negative crowding coefficient while high-income riders strictly dislike crowding.

while y_h=9.11 thousand RMB. The BLP estimates in Table 8 imply $(\alpha_\ell, \beta_\ell)$=(-0.213,-0.002) for the low-income group and (α_h, β_h)=(-0.174,-0.029) for the high-income group.

Baseline ridership follows directly from the data. The ridership-weighted average crowding density is 4.03 persons per square meter, car area is 53.2 m^2, and a train has J=6 cars, yielding an equilibrium ridership of N_{eq}=4.03×53.2×6=1,286. Given a ridership-weighted subway mode share of 0.43 (constructed from the BHTS imputation on our working sample), the market size is M=1,286/0.43≈2,983. Ridership-weighted average in-train travel time of 31.83 minutes gives L=31.83/(53.2×6)=0.100, and the ridership-weighted baseline fare is P_{base}=5.08 RMB. The mean utility A is calibrated so that the private equilibrium at P_{base} reproduces N_{eq}, which yields A=1.77. Table 9 reports parameter values used in the welfare and counterfactual exercises.

Table 9: Parameter Values Used in the Counterfactual Simulations

Parameter	Value	Description
Structural parameters (from Table 8)		
α_0	−0.179	Price coefficient, intercept
α_1	+0.013	Price coefficient, income slope
β_0	−0.026	Crowding coefficient, intercept
β_1	−0.009	Crowding coefficient, income slope
ρ_0	−0.017	Rescheduling coefficient, intercept
ρ_1	+0.002	Rescheduling coefficient, income slope
Group-level parameters, evaluated at demeaned incomes		
α_ℓ, α_h	−0.213, −0.174	Price sensitivity by group
β_ℓ, β_h	−0.002, −0.029	Crowding disutility by group
Market and calibration		
$y_\ell, y_h, \bar{y}$	6.19, 9.11, 8.77 (k RMB)	Group incomes (p41/p59) and sample mean
N_{eq}	1,286	Baseline equilibrium ridership
M	2,983	Total market size
L	0.100	Congestion exposure (min/m^2/person)
P_{base}	5.08 RMB	Baseline fare
A	1.77	Calibrated mean utility
Two-class configuration		
J_B, J_S	3, 3	Business, Standard car count
λ	0.5	Nesting parameter

Notes: Structural parameters are from the random-coefficient logit estimates in Table 8. Group-level coefficients are evaluated at demeaned incomes $y_g - \bar{y}$ where $\bar{y}$ = 8.77 thousand RMB is the sample mean. Baseline ridership, market size, and congestion exposure are computed from the sample. The mean utility A is calibrated to reproduce N_{eq} at P_{base}.

For the two-class counterfactual we partition the six-car train into 3 Business and 3 Standard cars.

The symmetric split is a natural benchmark because $L_B = L_S = 2L$ and the two classes are identical apart from the fare. Riding in either class has the same mean utility A. Class-specific utilities are $V_{gB} = A + \alpha_g P_B + \beta_g L_B N_B$ and $V_{gS} = A + \alpha_g P_S + \beta_g L_S N_S$. Any crowding difference between the two classes is therefore driven entirely by self-selection in response to the fare gap, not by a mechanical capacity asymmetry or differences in other features.

Through the well-known independence of irrelevant alternatives property, a standard multinomial logit would mechanically inflate the metro share once a second metro option is introduced, even if prices are identical. We therefore use a nested logit with a metro nest, so that the inclusive value for group g is

$$\mathrm{IV}_g \;=\; \lambda \,\ln\!\Big(e^{V_{gB}/\lambda} + e^{V_{gS}/\lambda}\Big), \tag{16}$$

and the corresponding choice probabilities follow the standard nested-logit form with nesting parameter $\lambda\in(0,1]$. When λ=1, the model collapses to multinomial logit. When λ <1, Business and Standard classes are more correlated with each other than either is with the outside option, which attenuates the spurious variety bonus. We use λ=0.5 as our preferred value, at the lower end of the range typical for sub-modes within a transit nest in the discrete-choice transport literature, and examine sensitivity to $\lambda\in[0.3,0.8]$ in the appendix.

6.2 Private Demand, Social Demand, and the Externality Wedge

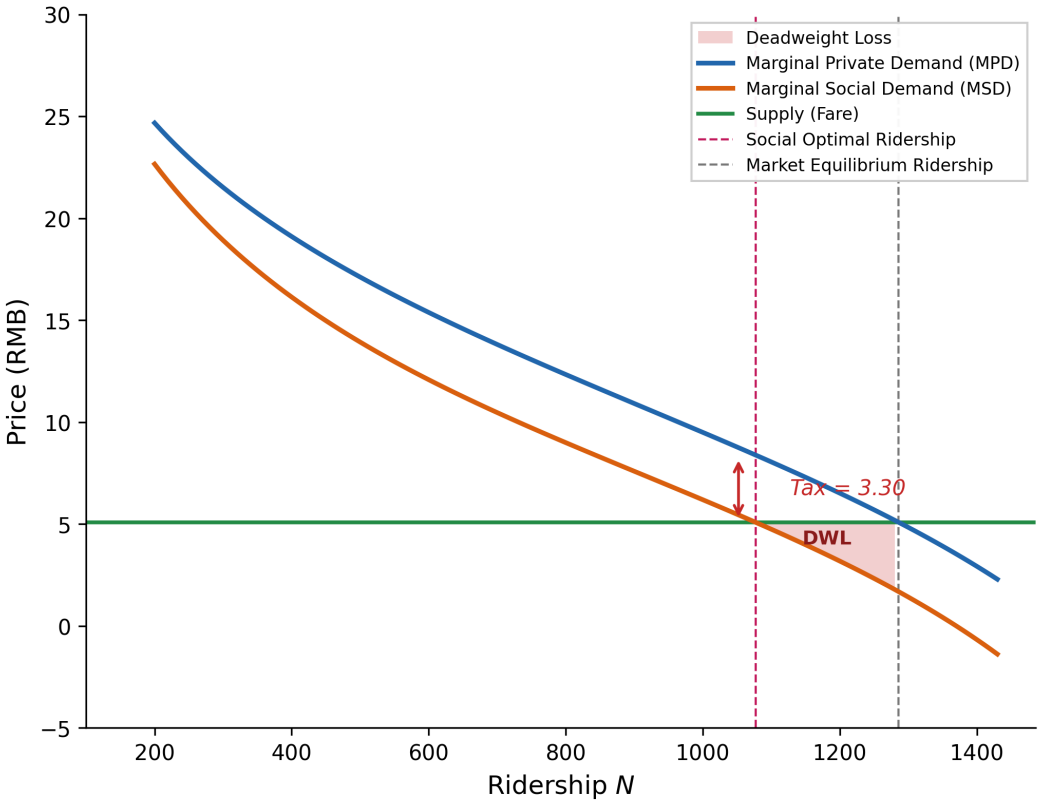
The welfare analysis builds on the distinction between marginal private demand and marginal social demand, illustrated in Figure 11. In the market equilibrium, the price equals each rider's marginal willingness to pay net of the crowding disutility the rider experiences personally. Each additional passenger, however, also imposes crowding on every other rider, a cost the private choice does not internalize. The social planner accounts for this externality, and the vertical distance between private demand and social demand captures the marginal externality.

To construct the inverse private demand, fix any candidate aggregate ridership N. The price that rationalizes N as a private equilibrium must satisfy the adding-up condition $N = \sum_g M_g\, s_g(A + \alpha_g P + \beta_g L N)$, which implicitly defines the mapping $PD(N)$. Letting $N_g(P, N) = M_g\, s_g(A + \alpha_g P + \beta_g L N)$ denote group-specific ridership, the externality wedge between private and social demand can be written as

$$w(N) \;\equiv\; PD(N) - MSD(N) \;=\; L \sum_{g\in\{\ell,h\}} \frac{\beta_g}{\alpha_g}\, N_g\big(PD(N), N\big). \tag{17}$$

The intuition is straightforward. Each additional rider raises crowding linearly in L, and the monetized disutility per unit of crowding for group g is β_g/α_g since price is the numeraire. Summing these monetized effects across groups, weighted by the number of riders exposed, gives the aggregate marginal external cost.

Figure 11: Crowding Externality and the Optimal Crowding Tax



Notes: MPD denotes marginal private demand, MSD marginal social demand, and the vertical distance between them is the externality wedge. At the social optimum N^*, the Pigouvian crowding tax equals the wedge $w(N^*)$, and the deadweight loss from unpriced crowding is the shaded triangle between MPD and MSD over $[N^*, N_{\text{eq}}]$.

Given the operator's price P_{base}, the constrained social optimum solves

$$PD(N^*) - w(N^*) \;=\; P_{\text{base}}, \tag{18}$$

and the corresponding Pigouvian crowding tax is $T^* = w(N^*)$, so that riders face $P_{\text{cons}} = P_{\text{base}} + T^*$. At P_{cons}, individual choices under the competitive equilibrium achieves the social optimal levels.

6.3 Counterfactual Policies

We evaluate three policy instruments against the baseline: a Pigouvian crowding tax, capacity-rationing queuing, and a two-class pricing scheme. Each holds the operator's fare revenue constant at $P_{\text{base}} \times N_{\text{eq}}$=6,533 RMB to ensure comparability.

Under a Pigouvian crowding tax, the social optimum N^* is obtained by solving equation (18) by bisection on N. The tax is $T^* = w(N^*)$, and riders face $P_{\text{cons}} = P_{\text{base}} + T^*$. Tax revenue is $T^* N^*$.

Capacity-rationing queuing is a non-price instrument in which each train has a fixed boarding

capacity equal to N^* and excess demand produces a queue of length W in minutes. The wait enters utility through two channels: as a time cost priced at the group-specific value of time VOT_g, which we take as half of the wage rate, and as a schedule-disruption cost captured by the rescheduling coefficient ρ_g estimated in Section 5.3. The generalized price is $P_{\text{base}} + \text{VOT}_g \times W$, and utility includes an additional term $\rho_g \times W$. The equilibrium wait W^* clears the queue, $\sum_g M_g\, s_g(V_g(W^*)) = N^*$, and we solve by bisection on W. Queuing generates no transferable revenue, and the wait cost is an additional deadweight loss; users decide to join the queue with rational expectations about the equilibrium waiting time.

The two-class counterfactual maintains revenue neutrality $P_B N_B + P_S N_S = P_{\text{base}}\, N_{\text{eq}}$. For any given P_S, we find $P_B > P_S$ and solve the nested-logit fixed point in (N_B, N_S), and retain the pair that maximizes the variety-adjusted consumer surplus defined in Section 6.4. The numerical implementation is described in Appendix C.

6.4 Welfare Calculation

We now describe how consumer surplus and total welfare are computed for each of the four counterfactuals introduced above. Because the idiosyncratic taste shocks are type-I extreme value, consumer surplus in each group admits the standard log-sum representation. In the single-class model, which covers the baseline, the Pigouvian tax, and the queuing counterfactual,

$$CS_g(N, P) \;=\; -\frac{M_g}{\alpha_g}\,\ln\!\Big(1 + e^{A+\alpha_g P+\beta_g LN}\Big), \qquad CS = \sum_g CS_g. \tag{19}$$

The factor $1/|\alpha_g|$ converts utility into monetary units, and the crowding externality is already embedded in V_g. In the two-class nested logit, the log-sum argument is replaced by the inclusive value, giving $CS_g = -(M_g/\alpha_g)\ln(1 + e^{\text{IV}_g})$.

The nested-logit formula mechanically exceeds single-class consumer surplus even when $P_B = P_S$, because the inclusive value exceeds the deterministic utility by Jensen's inequality. This artificial variety bonus is a known feature of discrete-choice models, which tend to overstate the welfare gains from expanding the choice set when the additional alternatives do not correspond to genuine product differentiation. Because relabeling physically identical cars as Business or Standard does not constitute real variety, we remove this artifact. Let VB denote the consumer-surplus gain when we impose the revenue-neutral equal price, $P_B = P_S = P_{\text{eq}}$, relative to baseline. The variety-adjusted consumer surplus is

$$CS_{\text{VA}} \;=\; CS_{\text{total}} - \text{VB}, \tag{20}$$

which by construction equals baseline consumer surplus in the absence of price differentiation. We report CS_{VA} as the primary welfare measure for two-class counterfactuals. Total welfare is reported as $W = CS + \text{TaxRev}$, with CS_{VA} substituting for CS in the two-class case. Under revenue neutrality,

fare revenue is held fixed across counterfactuals.

6.5 External Road Congestion

Changes in subway ridership also affect surface-road congestion as displaced riders shift to other modes. Let $\text{MECC}_{\text{veh-km}}$ denote the marginal external cost of road congestion per vehicle-kilometer, $\bar{d}$ the mean trip distance in kilometers, and κ the average car occupancy, so that the passenger-level congestion cost in cars is $c_{\text{car}} = \text{MECC}_{\text{veh-km}}/\kappa$. Diversion shares to other modes differ across groups. We assume that low-income riders shift to $p_{\text{bus},\ell} = 0.5$, $p_{\text{car},\ell} = 0.25$, and other modes with probability 0.25, while high-income riders shift to $p_{\text{bus},h} = 0.15$, $p_{\text{car},h} = 0.75$, and other modes with probability 0.10, reflecting the much higher rate of car ownership among higher-income households. Following Gu et al. (2023), each bus per kilometer generates 0.4 car-equivalents of congestion, while riders shifting to other modes are assumed not to affect road traffic. The change in road-congestion cost relative to baseline is

$$\Delta\text{MECC} \;=\; c_{\text{car}}\,\bar{d}\,\sum_{g}\left(p_{\text{car},g} + 0.4\,p_{\text{bus},g}\right)\left(N_g^{\text{base}} - N_g^{\text{cf}}\right). \tag{21}$$

We report welfare both without and with MECC, as W and $W - \Delta\text{MECC}$.

6.6 Counterfactual Results

Table 10 summarizes the comparison across policies. Panel A reports prices and ridership, and Panel B reports consumer surplus, revenue, and welfare.

The Pigouvian tax lowers ridership to the socially optimal level of N^*=1,077, a 16 percent reduction from baseline, through a crowding tax of T^*=3.30 RMB per trip. Consumer surplus falls from 10,885 to 7,671 RMB, but the tax generates 3,555 RMB in transferable revenue, so total welfare rises to 11,226 before accounting for road congestion. Once road-congestion costs from displaced riders are deducted, welfare reaches 10,737, still slightly below baseline. The incidence is sharply regressive. Low-income consumer surplus falls by 3,237 RMB, while high-income consumer surplus actually rises by 23 RMB, because higher-income riders gain nearly as much from the reduction in crowding as they lose from the higher fare. Even when tax revenue is redistributed by ridership share, the net welfare change is −707 RMB for low-income riders and +559 for high-income riders.

Capacity-rationing queuing also targets N^*, with an equilibrium wait of W^*=7.75 minutes, about 2.58 headways at h=3 minutes. Consumer surplus is 7,782 RMB. Because the instrument generates no revenue, total welfare stays close to consumer surplus and far below the Pigouvian benchmark. The queuing deadweight loss, computed as $\sum_g N_g\,W^*(\text{VOT}_g + |\rho_g/\alpha_g|)$, is approximately 3,434 RMB, almost exactly the revenue that the Pigouvian tax would collect. This equivalence is the classical

Table 10: Welfare Comparison across Counterfactual Scenarios

Panel A: Prices and Ridership

	P	tax	$W^\star$	N	N_ℓ	N_h	k
1. Single-class configuration							
1.1 Baseline (status quo)	5.08	–	–	1,286	1,132	154	4.03
1.2 Optimal crowding tax	8.38	3.30	–	1,077	919	158	3.37
1.3 Quantity control (queuing)	5.08	–	7.75	1,077	938	139	3.37
2. Two-class (3+3, $\lambda = 0.5$)							
2.1 Welfare maximization	–	–	–	1,524	1,235	289	4.77
Business	10.30	–	–	357	69	288	2.24
Standard	2.44	–	–	1,167	1,166	1	7.31

Panel B: Consumer Surplus, Revenue, and Welfare (1,000 RMB)

	Consumer Surplus			Tax	Δ Road Cong.	Welfare		Δ Welfare (w/ MECC)	
	CS	ΔCS_ℓ	ΔCS_h	Rev.	Exter.	w/o MECC	w/ MECC	ΔW_ℓ	ΔW_h
Baseline	10,885	0	0	0	–	10,885	10,885	0	0
Crowding tax	7,671	-3,237	23	3,555	489	11,226	10,737	-707	559
Queuing	7,782	-3,008	-94	0	524	7,782	7,258	-3,469	-158
Two-class	11,737	27	825	0	-820	11,737	12,556	272	1,399

Notes: Panel A reports the consumer price P (equal to $P_{\text{base}} + T^*$ under the crowding tax), tax rate, queue wait W^* (in minutes), ridership totals by group, and crowding density k (persons/m^2). Panel B reports consumer surplus and changes by group, tax revenue, the change in road congestion cost, and welfare with and without the MECC adjustment. The last two columns report group-level welfare changes inclusive of MECC: $\Delta W_g = \Delta CS_g + \text{TaxRev}_g - \text{MECC}_g$, so $\Delta W_\ell + \Delta W_h$ equals the change in welfare with MECC relative to baseline. The queuing scenario targets the social optimum $N^* = 1{,}077$. Two-class pricing uses the variety-adjusted consumer surplus $CS_{\text{VA}} = CS_{\text{total}} - \text{VB}$. The revenue-neutrality target is $P_{\text{base}} \times N_{\text{eq}} = 6{,}533$ RMB. All counterfactuals use $\lambda = 0.5$ and the 3+3 car allocation.

result from the queuing literature: rationing by time wastes as deadweight loss what rationing by price collects as transferable revenue.[14]

Two-class pricing produces the highest welfare among the three instruments. Revenue neutrality at λ=0.5 implies optimal fares of P_B^*=10.30 and P_S^*=2.44 RMB. Total ridership rises to 1,524, an 18.5 percent increase over baseline, as the discounted Standard fare attracts new riders from the outside option. Variety-adjusted consumer surplus reaches 11,737 RMB, a +7.82 percent gain. The ridership expansion also reduces road congestion, adding 820 RMB in MECC savings. Variety-adjusted welfare including MECC is 12,556 RMB, 15.4 percent above baseline. Distributionally, both groups gain on net once road-congestion benefits are redistributed by group: ΔW_ℓ = +272 and ΔW_h = +1,399. Two-class pricing is the only instrument under which no group is worse off.

Three lessons emerge from the comparison. First, transferable revenue is a first-order source of

[14]We omit a stand-alone availability-lottery scenario. As a non-price instrument it is observationally close to capacity-rationing queuing, and it would be implemented as ex-ante random access rather than realized waiting, which is unrealistic as a real-world subway policy.

welfare for any policy that reduces ridership. The Pigouvian tax dominates queuing purely because its revenue can be redeployed. Second, capacity-rationing queuing yields consumer surplus close to the Pigouvian level but very different total welfare, restating the classical point that rationing by time is allocatively similar to rationing by price but inferior because of the wasted transfer. Third, service differentiation combined with revenue-neutral pricing exploits the steep income gradient in crowding WTP documented in Section 5.3. High-income riders self-select into less crowded Business cars at a premium that cross-subsidizes lower Standard fares. Both groups gain on consumer surplus once road-congestion spillovers are included, so two-class pricing achieves a Pareto improvement.

7 Conclusions

This paper provides a revealed-preference estimate of subway passengers' willingness to pay to avoid crowding. Combining smartcard data from the Beijing Subway with an early-bird discount that creates plausibly exogenous variation in fares across departure times, we estimate that the marginal WTP to reduce in-train crowding by one passenger per square meter is about 0.06 RMB per minute. Aggregated over the average trip, this implies a crowding externality that exceeds the fare itself, suggesting that the unpriced congestion cost is a first-order distortion in subway pricing. The willingness to pay rises sharply with income while price sensitivity falls, so the average estimate masks substantial heterogeneity in how passengers value less crowded service. Passengers also dislike deviating from their preferred arrival time, with an MWTP of about 0.10 RMB per minute of schedule deviation. Unlike crowding, this rescheduling cost is essentially flat across the income distribution, indicating that high- and low-income riders place similar value on arriving on time even as they differ markedly in how much they will pay to avoid a crowded car.

Building on these estimates, we evaluate three policies that target the social optimum. A Pigouvian crowding tax raises welfare relative to the unpriced equilibrium, but the gain is driven entirely by transferable revenue, and the incidence is sharply regressive. Capacity-rationing queuing produces consumer surplus close to the Pigouvian level but underperforms the tax because it generates no revenue, with the queue's wait cost almost exactly equal to the foregone fiscal transfer. A two-class configuration combined with revenue-neutral pricing dominates the alternatives. Higher-income passengers self-select into a more expensive but less crowded Business car, the discounted Standard fare draws new riders from the outside option, and total welfare rises by roughly 15 percent above baseline once road-congestion spillovers are included. Both income groups are strictly better off, so two-class pricing is the only instrument under which no group bears a welfare loss.

Methodologically, the paper contributes to the literature on transit demand and congestion pricing along three dimensions. We use granular smartcard data to construct trip-level measures of in-train crowding from imputed routes, develop a novel instrumental variable based on overlapping trips between unrelated origin-destination pairs that addresses the endogeneity of crowding,

and combine demand estimation with a calibrated welfare model that quantifies both within-system surplus and external road-congestion effects. We show that the regressive incidence of a textbook Pigouvian tax can be overturned by a simple service-differentiation policy that exploits the income gradient in crowding valuations. From a policy standpoint, our results suggest that operators of crowded transit systems can achieve large efficiency gains and a Pareto improvement without raising average fares, provided that fare differentiation is accompanied by genuine differentiation in the riding experience.

References

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

_ , Andre De Palma, and Robin Lindsey, "A structural model of peak-period congestion: A traffic bottleneck with elastic demand," *The American Economic Review*, 1993, pp. 161–179.

Beijing Municipal Institute of City Planning and Design, "Trip-level Data in the Beijing Subway," 2015.

Beijing Transport Institute, "Household Travel Surveys," 2015.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Berry, Steven T, "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 1994, pp. 242–262.

Chu, Xuehao, "Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach," *Journal of Urban Economics*, 1995, *37* (3), 324–343.

Cook, Cody and Pearl Li, "Lexus Lanes or Corolla Lanes? Sorting and Equity in Expressway Pricing," *Working Paper*, 2025.

Coulombel, Nicolas and Guillaume Monchambert, "Diseconomies of scale and subsidies in urban public transportation," *Journal of Public Economics*, 2023, *223*, 104903.

Daganzo, Carlos F, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation research part B: methodological*, 1994, *28* (4), 269–287.

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

Douglas, Neil and George Karpouzis, "Estimating the passenger cost of train overcrowding," in "29th Australian Transport Research Forum" 2006, pp. 1–8.

Gu, Yizhen, Qu Tang, Yacan Wang, and Ben Zou, "Fare structure and the demand for public transit," *Work Pap*, 2023.

Hahn, Robert W, Robert D Metcalfe, and Eddy Tam, "Welfare Estimates of Shifting Peak Travel," Technical Report, National Bureau of Economic Research 2023.

Hall, Jonathan D, "Inframarginal travelers and transportation policy," *International Economic Review*, 2024, *65* (3), 1519–1550.

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

Henderson, J Vernon, "Road congestion: a reconsideration of pricing theory," *Journal of Urban Economics*, 1974, *1* (3), 346–365.

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

Hortaçsu, Ali, Olivia R Natan, Hayden Parsley, Timothy Schwieg, and Kevin R Williams, "Organizational structure and pricing: Evidence from a large us airline," *The Quarterly Journal of Economics*, 2024, *139* (2), 1149–1199.

Kreindler, Gabriel, "Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications," *Econometrica*, 2024, *92* (4), 1233–1268.

Li, Shanjun, "Better Lucky than Rich? Welfare Analysis of Automobile License Allocations in Beijing and Shanghai," *Review of Economic Studies*, 2018, *85* (4), 2354–2383.

Li, Zheng and David A Hensher, "Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal," *Transport Policy*, 2011, *18* (6), 880–887.

Lighthill, Michael James and Gerald Beresford Whitham, "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*, 1955, *229* (1178), 317–345.

Lu, Hui, Tony Fowkes, and Mark Wardman, "Amending the incentive for strategic bias in stated preference studies: case study in users' valuation of rolling stock," *Transportation research record*, 2008, *2049* (1), 128–135.

Parry, Ian W H and Kenneth A Small, "Should urban transit subsidies be reduced?," *American Economic Review*, 2009, *99* (3), 700–724.

Richards, Paul I, "Shock waves on the highway," *Operations research*, 1956, *4* (1), 42–51.

Singh, Jyotsna, Gonçalo Homem de Almeida Correia, Bert van Wee, and Natalia Barbour, "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*, 2023, *170*, 103628.

Vickrey, William S, "Congestion theory and transport investment," *The American economic review*, 1969, *59* (2), 251–260.

Wardman, Mark and Gerard Whelan, "Twenty years of rail crowding valuation studies: evidence and lessons from British experience," *Transport reviews*, 2011, *31* (3), 379–398.

Weitzman, Martin L., "Prices vs. Quantities," *Review of Economic Studies*, 1974, *41* (4), 477–491.

Yap, Menno, Oded Cats, and Bart Van Arem, "Crowding valuation in urban tram and bus transportation based on smart card data," *Transportmetrica A: Transport Science*, 2020, *16* (1), 23–42.

Crowding
Online Appendix (Not for Publication)

Yizhen Gu Qu Tang Kai Wu Ben Zou

Table of Contents

A Data Sources

This appendix summarizes the data sources used in the paper.

Beijing Subway smartcard records. The primary data source is the trip-level records from the Beijing Subway's automated fare collection (AFC) system (Beijing Municipal Institute of City Planning and Design, 2015). Each record captures the tap-in and tap-out stations and timestamps for a single trip, and trips can be linked across days through an anonymized card identifier. We estimate that between 90 and 95 percent of subway trips were paid by smartcard around the time of the EBD. The main sample consists of the universe of smartcard trips on 21 workdays in five non-consecutive weeks in 2016, covering trips originating from the 16 EBD stations and the 16 control stations during the 6:30 to 8:30 AM window.

Subway network and operating information. We geocode all subway stations and collect the track distance and fare for each station pair from the Beijing Subway's official website. We also collect train frequencies by line and station, the types of passenger cars, and the number of cars per train. These inputs are used to compute the total floor area available on each train segment at each moment and to determine the scheduled in-train travel time between adjacent stations. We treat seating and standing areas identically.

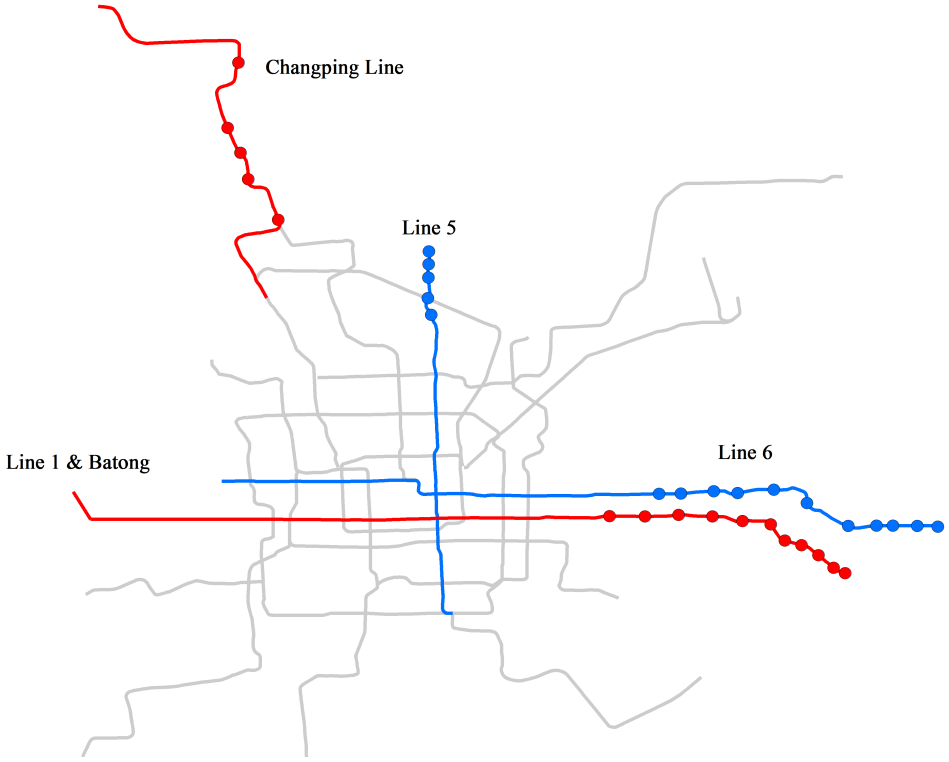
2015 Beijing Household Travel Survey. The Beijing Household Travel Survey (Beijing Transport Institute, 2015) is a large representative survey of approximately 100,000 households conducted by the Beijing Transport Institute. It records each respondent's home and work locations, daily travel details, and transportation mode choices. We use the BHTS to fit the multinomial logit mode-choice model used in defining market sizes and the outside option.

Baidu Maps commuting flow data. We obtain grid-level commuting flow data from Baidu Maps, a leading digital map provider in China. The data track usual nighttime (home) and daytime (work) locations of roughly 7.6 million devices across about 15,000 residential grids with an edge of approximately 700 meters. For each home-work grid pair, the data report the share of commuters in each of five monthly income brackets. Income is imputed by Baidu's internal machine-learning algorithm from device characteristics, installed applications, and location patterns. We use the Baidu data to construct OD-specific income distributions of potential subway riders.

Auxiliary sources. We collect supplementary information on the 16 EBD stations and their parallel control stations from the Beijing Subway's public announcements. Official annual ridership statistics used to benchmark smartcard coverage are from the Beijing Municipal Commission of Transport.

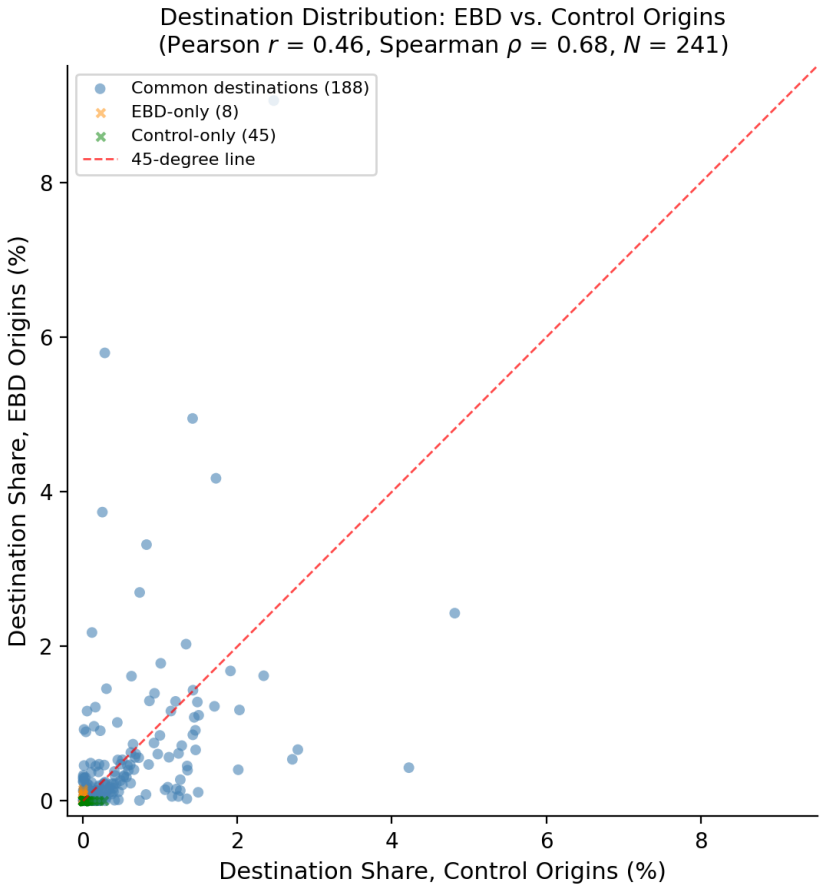
B Additional Empirical Results

Figure B.1: Beijing Subway Map with EBD and Control Stations



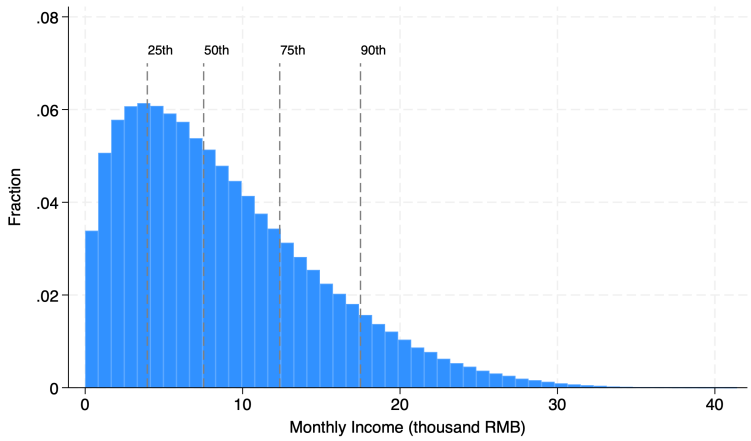
Notes: This figure shows the Beijing Subway network with the 16 early-bird discount (EBD) stations marked in red and the 16 control stations marked in blue. EBD stations are located on the Batong Line and the Changping Line. Control stations are on Line 5 and Line 6.

Figure B.2: Destination Distribution of EBD and Control Station Riders



Notes: Each point represents a destination station. The horizontal axis is the share of all control-origin riders going to that destination; the vertical axis is the corresponding share for EBD-origin riders. Blue dots denote the 188 destinations served by both groups; orange and green crosses denote the 8 EBD-only and 45 control-only destinations, respectively. The red dashed line is the 45-degree line. Pearson $r = 0.46$, Spearman $\rho = 0.68$ ($N = 241$).

Figure B.3: Simulated Monthly Income of Subway Riders (Thousand RMB)



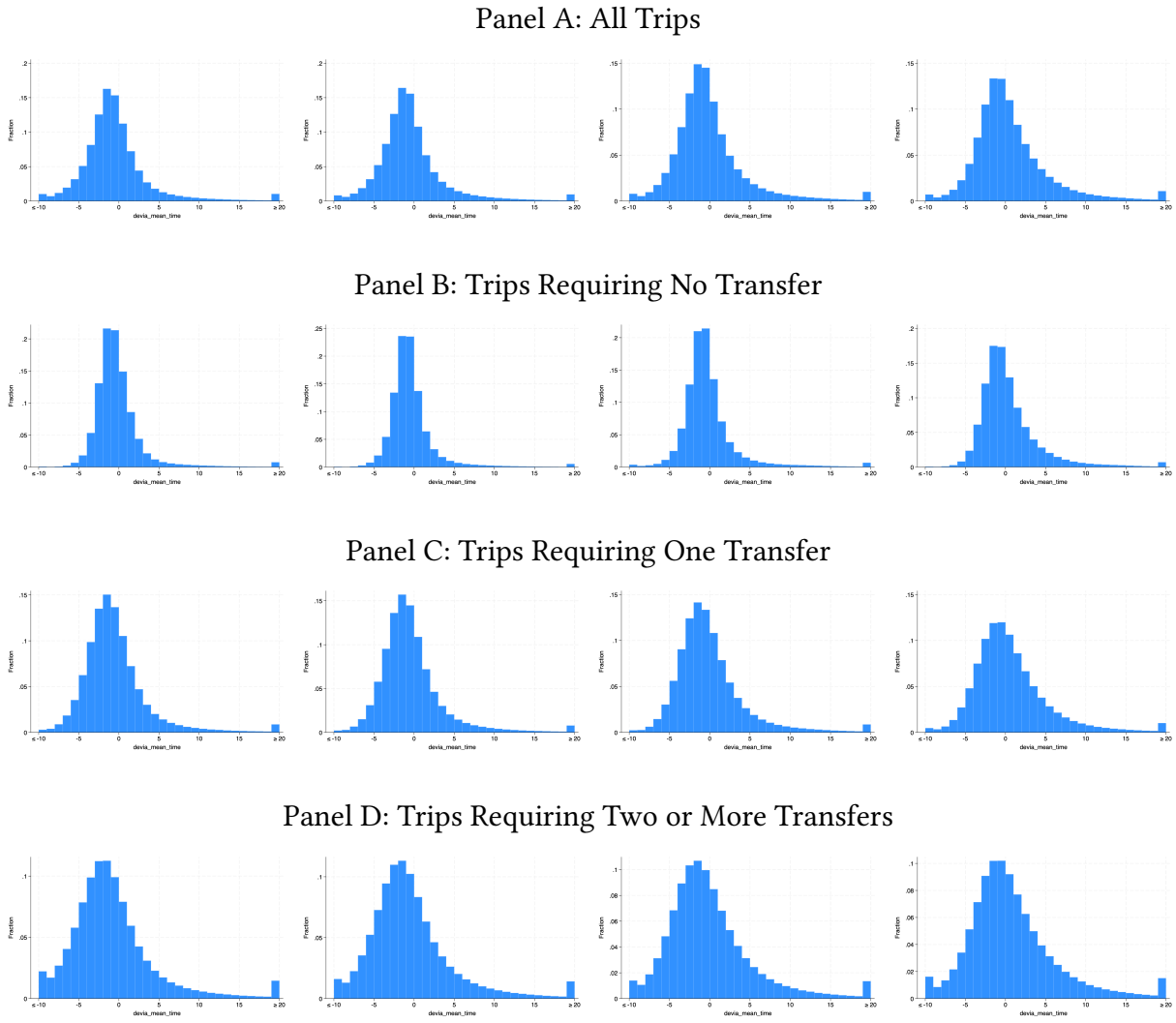
Notes: Distribution of simulated monthly incomes (in thousands of RMB) drawn from OD-specific beta distributions fitted to the Baidu commuting data. The weighted mean income is 8.77 thousand RMB.

Table B.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

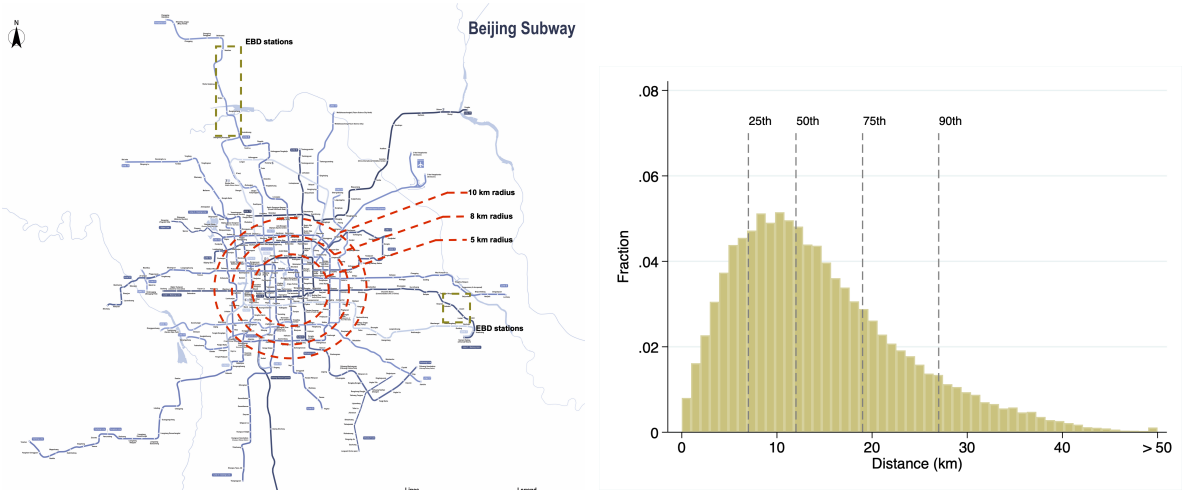
Notes: This table reports estimated in-station times from the iterative calibration procedure described in Section 3.3.1. Entry time is the duration from tapping in at the origin station to boarding the train. Exit time is the duration from alighting to tapping out, with the mean normalized to 2 minutes. Transfer time is the duration at each transfer station. The pooled transfer row reports the average across all transfers; subsequent rows break out estimates by transfer order (first, second, etc.). N denotes the number of trips used in estimation.

Figure B.4: Distribution of Travel Time Deviations by Crowding Quartile and Number of Transfers



Notes: Each row shows the distribution of deviations from the OD-specific mean travel time (in minutes) for a given transfer category, split by quartiles of in-train crowding density (least crowded on the left, most crowded on the right). Panel A includes all trips; Panel B includes trips without transfers; Panel C includes trips with one transfer; Panel D includes trips with two or more transfers. The distributions remain similar across crowding quartiles within each transfer category, indicating that the right tail grows with the number of transfers rather than with crowding levels.

Figure B.5: ACP Geography and Distance Distribution

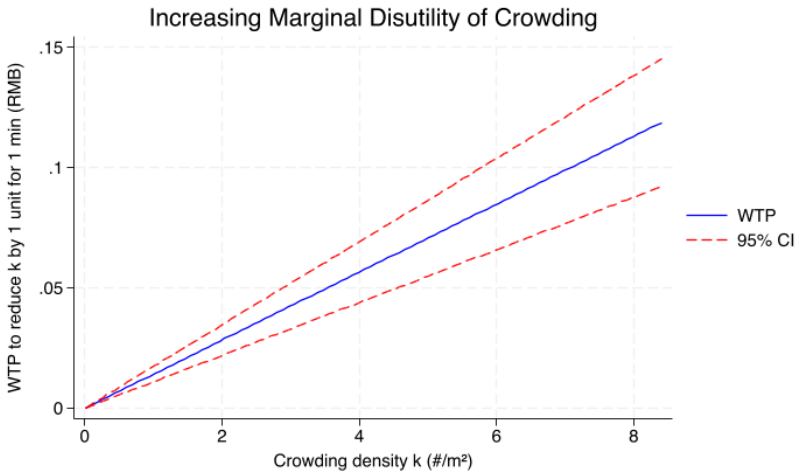


(a) Beijing Subway Network and Radius Rings

(b) Distance from ACP Destination D' to Focal Destination D (Infrequent Riders)

Notes: Panel (a) shows the Beijing Subway network. EBD stations, located on the Changping Line (north) and the Batong Line (southeast), are highlighted in dashed boxes at the far suburban ends of the network. Red dashed circles mark 5-km, 8-km, and 10-km radii from the city center. The vast majority of the network lies beyond the 8-km ring, illustrating that the exclusion zone around a focal destination D removes only a small fraction of potential ACPs. Panel (b) shows the distribution of distances between each infrequent ACP's destination D' and the focal passenger's destination D. Dashed vertical lines mark the 25th, 50th, 75th, and 90th percentiles. The median ACP destination is approximately 11 km from D, well beyond the 8-km exclusion threshold used in the preferred IV specification. Infrequent riders are defined as those taking ≤ 3 trips per week.

Figure B.6: Increasing Marginal Disutility of Crowding



Notes: WTP for crowding reduction by crowding density level, estimated from a non-linear specification with Crowd^2 and Crowd^3 terms, using the multi-IV instrument set. The solid line plots the estimated MWTP as a function of crowding density; dashed lines show the 95% confidence interval. The increasing pattern confirms convex crowding disutility.

Table B.2: GMM Estimates with ACP as Instrumental Variable

	GMM: Single IV (ACP)		GMM: Multiple IV (ACP)	
	1st Stage	2nd Stage	1st Stage	2nd Stage
	(1)	(2)	(3)	(4)
Crowding		−0.0186		−0.0085
		(0.0020)		(0.0011)
Price	−5.0395	−0.3214		−0.1794
	(0.4364)	(0.0200)		(0.0094)
Rescheduling	0.0429	−0.0275	−0.1369	−0.0248
	(0.0307)	(0.0014)	(0.0240)	(0.0010)
ACP	180.0126		181.2738	
	(8.3763)		(8.1651)	
Other IVs			✓	
Wald *F*-Test		461.9		288.3
Observations	295,400	295,400	295,400	295,400
OD-Day FE	X	X	X	X
Time-Day FE	X	X	X	X
Implied Willingness to Pay				
MWTP for Crowding Reduction		0.0580		0.0474
		(0.0056)		(0.0061)
MWTP for Resched. Avoidance		0.0856		0.1377
		(0.0072)		(0.0091)

Notes: This table reports GMM estimates using the number of accidental companion passengers (ACP) rather than accidental companion time (ACT) as the instrumental variable for crowding. Column (1) reports the first-stage regression of in-train crowding on ACP; Column (2) reports the second-stage demand estimates with ln(share) as the dependent variable. Both columns include OD-day and time-day fixed effects. The sample is restricted to frequent riders. Standard errors in parentheses, clustered at the OD level. MWTP computed as the ratio of the relevant coefficient to the price coefficient.

Table B.3: Robustness: In-Station Crowding

	Single IV (ACT)		Multiple IV (ACT + EBD policy)	
	Baseline	+ In-Station (control)	Baseline	+ In-Station (endogenous)
	(1)	(2)	(3)	(4)
In-train crowding	−0.0201	−0.0293	−0.0094	−0.0078
	(0.0019)	(0.0026)	(0.0011)	(0.0013)
Price	−0.3280	−0.5855	−0.1821	−0.1618
	(0.0196)	(0.0323)	(0.0094)	(0.0097)
Rescheduling	−0.0274	−0.0291	−0.0250	−0.0242
	(0.0014)	(0.0014)	(0.0010)	(0.0011)
In-station crowding		0.0403		−0.0077
		(0.0030)		(0.0023)
Mean in-train crowding (person·min/m^2)	127.09	127.09	127.09	127.09
Mean in-station crowding (thousand person·min)	14.31	14.31	14.31	14.31
Observations	295,400	295,400	295,400	295,400
Wald *F*-stat	580.21	424.40	284.82	470.58
WTP (crowd)	0.0613	0.0500	0.0516	0.0482
	(0.0055)	(0.0029)	(0.0061)	(0.0075)
OD-Day FE	X	X	X	X
Time-Day FE	X	X	X	X

Notes: Dependent variable is $\ln(S_{jdt})$ from the linear estimating equation (4), estimated by GMM. Sample: frequent riders. Standard errors clustered at the OD level. The in-station crowding regressor InStn_{jdt} is the total in-station ridership-time per product (the sum across all riders in the product cell of their individual entry, transfer, and exit times), expressed in thousands of person-minutes. Columns (1) and (3) are the single-IV (ACT only) and multi-IV baselines without InStn_{jdt}. Column (2) adds InStn_{jdt} as an uninstrumented control under the single-IV instrument set; this specification is mis-specified because InStn_{jdt} is itself endogenous, and the positive coefficient on InStn_{jdt} reflects that endogeneity. Column (4) treats InStn_{jdt} as a second endogenous regressor, jointly instrumented with in-train crowding by the multi-IV set: ACT (the infrequent-companion-time instrument, restricted to companions whose destinations lie at least 8 km from the focal D) plus the four EBD policy interactions treat $\times\, 1\{t < 7\text{am}\}$ and treat $\times\, t^k$ for $k = 1, 2, 3$. The two "Mean" rows report regression-sample averages of in-train crowding density (person·min/m^2) and total in-station ridership-time (thousand person·min); both rows are flat across columns since the estimation sample is the same. WTP for in-train crowding reduction is $-\hat{\beta}_{\text{crowd}}/\hat{\alpha}$, reported with delta-method standard errors.

Table B.4: First Stage Estimation for Crowding (Multi-IV)

	(1)	(2)	(3)
ACT	11.2355	11.1407	10.6604
	(0.4516)	(0.4645)	(0.4360)
Treat × Pre-7am	51.7864	51.8486	34.5995
	(3.5895)	(3.5914)	(3.6339)
Treat × Trend	544.1930	543.4448	602.2188
	(36.3108)	(36.3020)	(40.2235)
Treat × Trend^2	-18.0093	-17.9853	-19.7376
	(1.2180)	(1.2179)	(1.3458)
Treat × Trend^3	0.1966	0.1963	0.2133
	(0.0136)	(0.0136)	(0.0149)
Price	29.3654	29.3990	18.8183
	(2.1952)	(2.1966)	(2.1249)
Rescheduling	-0.1323	-0.1360	-0.1368
	(0.0241)	(0.0243)	(0.0207)
Observations	295,400	295,400	295,400
OD-Day FE	X	X	X
Time FE	X		
Time-Day FE		X	
Time-Destin Line FE			X

Notes: This table reports the first stage of the multi-IV GMM specification for the frequent riders sample. The dependent variable is predicted in-train crowding (person·min/m^2). Instruments are ACT from infrequent riders with $|D - D'| \geq 8$ km, plus four EBD policy interactions. Standard errors in parentheses, clustered at the OD level.

C Details of Welfare and Counterfactual Analysis

This appendix collects the numerical implementation details for the counterfactual exercises of Section 6 and reports sensitivity analyses with respect to the car allocation between Business and Standard classes.

C.1 Numerical Implementation

Our counterfactual computations rely on three numerical primitives: one-dimensional bisection for root finding, damped fixed-point iteration for the two-class ridership vector, and a scan-bracket-bisect procedure to enforce revenue neutrality. We summarize each here.

One-dimensional roots are computed by bisection. Given a continuous function F and an interval $[a, b]$ with $F(a)F(b) < 0$, we iterate $m = (a+b)/2$ and retain the sub-interval that preserves the sign change until $|F(m)|$ falls below a tolerance set between 10^{-9} and 10^{-12} depending on the application. We use this routine in three places: to invert the private demand mapping, to locate the Pigouvian social optimum, and to compute the queuing equilibrium wait.

The inverse private demand $PD(N)$ is evaluated as follows. For each candidate N, we bisect on P to solve $f(P; N) = \sum_g M_g\, s_g(A + \alpha_g P + \beta_g LN) - N = 0$. The function $f(\cdot; N)$ is continuous in P and strictly decreasing when $\alpha_g < 0$, which guarantees a unique root in a sufficiently wide bracket. The baseline calibration of A uses the same routine in reverse: we bisect on A so that the implicit solution of $PD(N) = P_{\text{base}}$ equals N_{eq}.

The two-class ridership vector (N_B, N_S) at prices (P_B, P_S) is obtained by damped fixed-point iteration. Starting from a feasible initial guess, we update

$$N_B^{(k+1)} \;\leftarrow\; (1-\eta)N_B^{(k)} + \eta \sum_g M_g\, p_{gB}\big(V_{gB}(N_B^{(k)}), V_{gS}(N_S^{(k)})\big),$$

and analogously for $N_S^{(k+1)}$, with damping parameter $\eta \in [0.5, 0.7]$. Damping keeps the iterations stable when the nested-logit probabilities have steep local slopes in the class shares. Convergence is declared when both components change by less than 10^{-8} between successive iterations.

Revenue neutrality for the two-class counterfactual requires simultaneous control over P_B, P_S, and the equilibrium ridership vector. For each candidate P_S, we scan P_B upward from P_S to bracket every sign change in $R(P_B, P_S) - R_{\text{base}}$, where $R_{\text{base}} = P_{\text{base}} \times N_{\text{eq}} = 6{,}533$ RMB is the revenue target. Within each bracket we bisect to find a revenue-neutral P_B, and among all revenue-neutral pairs (P_B, P_S) we retain the one that maximizes the variety-adjusted consumer surplus. A final one-dimensional refinement over P_S yields the reported optimum.

Sensitivity to λ is reported in Table C.1. As λ rises, the variety bonus grows, and CS_{VA} falls toward baseline consumer surplus. At $\lambda = 0.8$, CS_{VA} is only +0.77 percent above baseline because revenue

neutrality forces $P_B \approx P_S$: the nested logit with a high λ penalizes unequal prices too strongly for self-selection to be sustained. At $\lambda = 0.3$, the gain reaches 15.6 percent, though such low nesting parameters are at the extreme end of the plausible range. The preferred value $\lambda = 0.5$ yields a 7.82 percent variety-adjusted welfare gain and is a conservative choice within the typical range.

Table C.1: Two-Class Pricing: Sensitivity to the Nesting Parameter λ

λ	P_B	P_S	N	VB	CS_{VA}	$\Delta\text{CS}_{\text{VA}}$ (%)
0.3	10.73	2.51	1538	1434	12579	15.56
0.4	10.63	2.48	1532	1918	12133	11.47
0.5	10.30	2.44	1524	2405	11737	7.82
0.6	9.74	2.43	1514	2896	11403	4.76
0.7	8.95	2.46	1505	3390	11144	2.38
0.8	7.76	2.64	1495	3888	10968	0.77

Notes: All scenarios maintain revenue neutrality (R = 6,533 RMB) with a 3+3 Business/Standard car allocation. VB denotes the variety bonus at the revenue-neutral equal price, $CS_{\text{VA}} = CS_{\text{total}} - \text{VB}$ is the variety-adjusted consumer surplus, and ΔCS_{VA} is the percentage change relative to baseline consumer surplus (10,885 RMB).

C.2 Sensitivity to Car Allocation

We examine how the two-class results vary with the Business/Standard split, holding $\lambda = 0.5$ fixed and maintaining revenue neutrality. A clear threshold emerges from Table C.2. When the number of Standard cars exceeds the number of Business cars, as in the 1+5 and 2+4 splits, the welfare optimizer fails to generate meaningful price differentiation. The revenue-neutral welfare-maximizing prices collapse to $P_B \approx P_S$, the price differential is numerically indistinguishable from zero, and the variety-adjusted consumer surplus gain vanishes. The mechanism is direct. When the Standard class has excess capacity, per-rider crowding in Standard is already low at the revenue-neutral equal price, so riders have little incentive to pay a premium for Business class. The optimizer cannot sustain self-selection because the crowding differential between classes disappears mechanically.

Splits with $J_B \geq J_S$ produce substantial price differentiation. Under the current calibration, the preferred 3+3 allocation yields $P_B^\star = 10.30$, $P_S^\star = 2.44$, with a variety-adjusted consumer surplus gain of 7.82 percent (Table C.1, row $\lambda = 0.5$). Other splits in Table C.2 (1+5, 2+4, 4+2, 5+1) have not been re-run under the current calibration; the qualitative threshold remains: revenue-neutral two-class pricing is effective only when $J_B \geq J_S$. Below this threshold, revenue neutrality forces the two-class system to collapse to a single uniform fare.

Table C.2: Two-Class Pricing: Sensitivity to Car Allocation

	$J_S > J_B$			$J_B \geq J_S$	
	1+5	2+4	3+3	4+2	5+1
$P_B^\star$ (yuan)	4.46	4.43	6.98	5.77	4.80
$P_S^\star$ (yuan)	4.43	4.43	2.97	2.96	3.57
$P_B^\star - P_S^\star$ (yuan)	0.03	0.00	4.01	2.81	1.22
Total ridership (N)	1,506	1,508	1,543	1,560	1,523
Price discrim. gain (RMB)	−2.7	−0.6	108.1	231.1	58.0
$\Delta \text{CS}_{\text{VA}}$ (%)	−0.03	−0.01	+1.17	+2.50	+0.63

Notes: All scenarios maintain revenue neutrality (R = 6,533 RMB) with $\lambda = 0.5$ and $J_B + J_S = 6$. Price discrimination gain is the difference between CS at the welfare-maximizing revenue-neutral prices and CS at the revenue-neutral equal price. ΔCS_{VA} is the percentage change relative to baseline single-class consumer surplus (10,885 RMB; the value shown for the 3+3 row reflects the current calibration, while the other rows are pre-recalibration and have not been regenerated).