Crowding

Yizhen Gu (Peking University)   Qu Tang (Jinan University)
Kai Wu (UT Austin)   Ben Zou (Purdue)

ABFER 2026

Motivation

- Crowding is a common disamenity.
  - Ubiquitous; disutility due to discomfort.
  - *Negative externality* on other people.

Motivation

- Crowding is a common disamenity.
  - Ubiquitous; disutility due to discomfort.
  - *Negative externality* on other people.
- Q: What is the WTP to reduce crowding?

Motivation

- Crowding is a common disamenity.
  - Ubiquitous; disutility due to discomfort.
  - *Negative externality* on other people.
- Q: What is the WTP to reduce crowding?
- An important parameter in public good provision.

Motivation

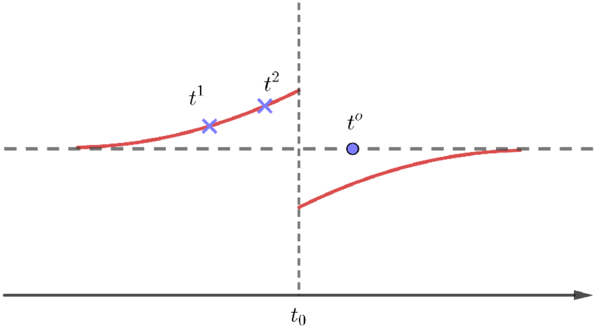
- Crowding is a common disamenity.
  - Ubiquitous; disutility due to discomfort.
  - *Negative externality* on other people.
- Q: What is the WTP to reduce crowding?
- An important parameter in public good provision.
- Empirical challenges
  - **Correlated demand shocks** (delicious food, exciting games, common work hours).
  - **Other attributes change with crowding** (speed, services, class status).

This Paper: Estimate MWTP to Avoid Crowding in Subway

- Empirical Setting: Crowded subway cars in morning peak hours in Beijing.
  - Unique advantage: crowding does **not** slow down the train!

This Paper: Estimate MWTP to Avoid Crowding in Subway

- Empirical Setting: Crowded subway cars in morning peak hours in Beijing.
  - Unique advantage: crowding does **not** slow down the train!
- Identification: Temporal price variation created by an early-bird discount (EBD).
  - Trade-offs btw **price**, **crowdedness**, and **deviation from the optimal schedule**.
  - Choice of departure time reveals WTP to avoid crowding.



Results Preview

- Estimate MWTP to avoid crowding
  - Avg in-train density on EBD lines, morning peak: ≈ 4 persons/m^2.
  - WTP to reduce one person/m^2 for one minute: ≈ 0.05–0.06 RMB.
  - Larger at higher crowding levels; larger for higher-income passengers.

Results Preview

- Estimate MWTP to avoid crowding
  - Avg in-train density on EBD lines, morning peak: ≈ 4 persons/m^2.
  - WTP to reduce one person/m^2 for one minute: $\approx$ 0.05–0.06 RMB.
  - Larger at higher crowding levels; larger for higher-income passengers.
- Optimal crowding tax
  - Crowding externality ≈ 6 RMB for a representative 32-min trip (fare ≈ 5 RMB).
  - Optimal crowding tax $T^\star \approx 3.3$ RMB; optimal ridership $\approx 16\%$ lower.

Results Preview

- Estimate MWTP to avoid crowding
  - Avg in-train density on EBD lines, morning peak: ≈ 4 persons/m^2.
  - WTP to reduce one person/m^2 for one minute: ≈ 0.05–0.06 RMB.
  - Larger at higher crowding levels; larger for higher-income passengers.
- Optimal crowding tax
  - Crowding externality ≈ 6 RMB for a representative 32-min trip (fare ≈ 5 RMB).
  - Optimal crowding tax $T^\star \approx 3.3$ RMB; optimal ridership $\approx 16\%$ lower.
- Alternative policies
  - Crowding tax is regressive; queuing dissipates revenue as deadweight loss.
  - Two-class configuration (revenue-neutral): total welfare $\approx 15\%$ above baseline; *both* high- and low-income groups gain.

Contributions

- Disutility from crowding in monetary terms via revealed preference
  - Most studies: stated preferences from surveys (Douglas & Karpouzis 2006; Wardman & Whelan 2011; Haywood & Koning 2015).
  - RP estimates via waiting or routing (Hörcher et al. 2017; Yap et al. 2020): WTP in *time*, not money.
  - Different from road congestion literature: discomfort vs. lost time (Vickrey 1969; Arnott, de Palma & Lindsey 1993; Parry & Small 2009).
- Optimal pricing and policy under crowding externality
  - First estimate of optimal crowding tax (cf. Parry & Small 2009; Coulombel & Monchambert 2023).
  - Two-class pricing induces self-selection and yields welfare gains (Cook & Li 2025 on HOT lanes).
- Price vs. quantity controls debate (Weitzman 1974; Li 2018).
- Measurements from large-scale transit card data
  - Real-time location, crowding imputations; ACP-based IV for crowding (cf. Hortacsu et al. 2024); coworker-based optimal arrival time.

Demand Model

Standard Logit Model

Commuter i on OD pair j, day d, chooses tap-in time bin t:

$$U_{ijdt} = V_{ijdt} + \varepsilon_{ijdt}, \quad \varepsilon_{ijdt} \sim \text{i.i.d. Type-I EV}$$

– Logit framework: commuters choose departure time to maximize utility.

Standard Logit Model

Commuter i on OD pair j, day d, chooses tap-in time bin t:

$$V_{ijdt} = \alpha_i P_{jdt} + \cdots$$

– Price: fare varies by time t within OD j due to EBD discount.

Standard Logit Model

Commuter i on OD pair j, day d, chooses tap-in time bin t:

$$V_{ijdt} = \alpha_i P_{jdt} + \beta_i \, \mathbb{E}_d[\text{Crowd}_{jdt}] + \cdots$$

– Crowding: expected in-train crowding experience (person·min/m^2) over the trip.

Standard Logit Model

Commuter i on OD pair j, day d, chooses tap-in time bin t:

$$V_{ijdt} = \alpha_i P_{jdt} + \beta_i \, \mathbb{E}_d[\text{Crowd}_{jdt}] + \rho_i \left| t + \mathbb{E}_d[\text{TT}_{jdt}] - t_{ijd}^{OA} \right| + \cdots$$

– Rescheduling cost: deviation of arrival time from individual optimal arrival t_{ijd}^{OA}.

Standard Logit Model

Commuter i on OD pair j, day d, chooses tap-in time bin t:

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

– Fixed effects & demand shocks:

- κ_{jd} (OD-day)
- ϕ_t (time-bin)
- ξ_{jdt} (demand shock)

Standard Logit Model

Commuter i on OD pair j, day d, chooses tap-in time bin t:

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

Estimating equation (Berry 1994):

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

Log share ratio: invert market shares to recover mean utilities.

Random-Coefficient Model (Income Heterogeneity)

- Income-dependent preference coefficients:

$$\alpha_i = \alpha_0 + \alpha_1 I_i, \quad \beta_i = \beta_0 + \beta_1 I_i, \quad \rho_i = \rho_0 + \rho_1 I_i$$

Price sensitivity, crowding disutility, and rescheduling cost all vary with income I_i (recentered at $\bar{y}$).

- Decompose utility into mean δ_{jdt} and individual deviation μ_{ijdt}:

$$V_{ijdt} = \underbrace{\alpha_0 P + \beta_0 \widehat{\text{Crowd}} + \rho_0 |\cdots| + \kappa_{jd} + \phi_t + \xi_{jdt}}_{\delta_{jdt}} + \underbrace{\alpha_1 I_i P + \beta_1 I_i \widehat{\text{Crowd}} + \rho_1 I_i |\cdots|}_{\mu_{ijdt}}$$

Key Challenges to Estimation

- Measurement & Price exogeneity
  - **Crowding:** Optimal route & train assignment.
  - **Optimal arrival time:** imputed from "coworkers".
  - **Price:** exogenous variation from the EBD (30% off before 7 AM at 16 stations); matched control stations.

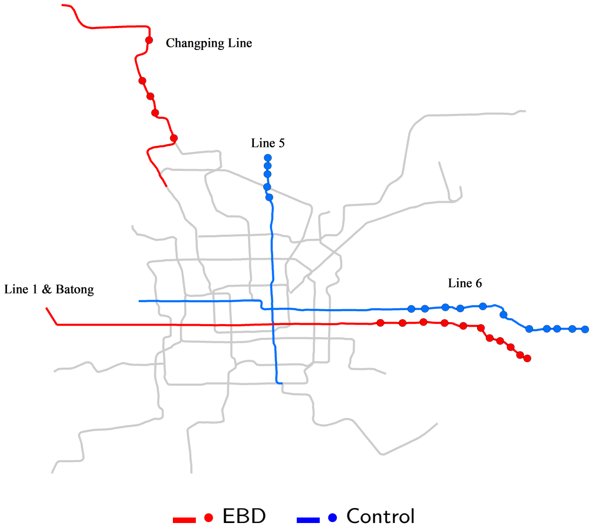
Key Challenges to Estimation

- Measurement & Price exogeneity
  - **Crowding:** Optimal route & train assignment.
  - **Optimal arrival time:** imputed from "coworkers".
  - **Price:** exogenous variation from the EBD (30% off before 7 AM at 16 stations); matched control stations.
- Crowding endogeneity
  - Positively corr. w/ demand shocks ξ_{jdt}; OLS biases $\hat{\beta}$ upward.
  - IV from Accidental Companion Passengers (ACP): but overlap in trip (thus contributing to crowding).

Background & Data

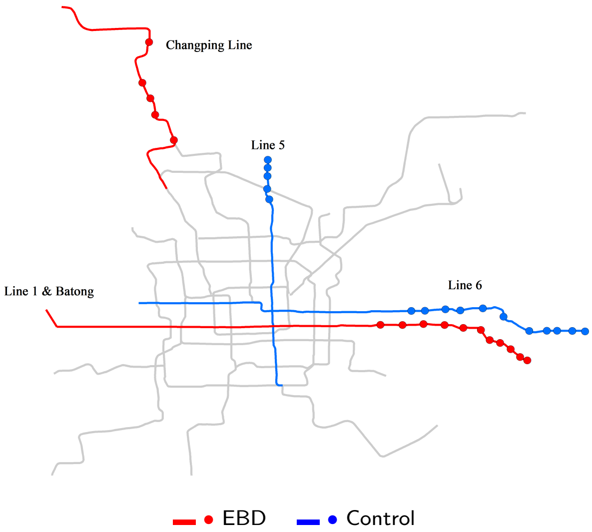
Background

- The Beijing Subway
  - 19 lines, more than 340 stations as of 2016.
  - ≈9 million daily trips.
  - Fare: 3–10 RMB, distance-indexed.



Background

- The Beijing Subway
  - 19 lines, more than 340 stations as of 2016.
  - ≈9 million daily trips.
  - Fare: 3–10 RMB, distance-indexed.
- Early Bird Discount (EBD)
  - 30% off before 7 AM at 16 EBD stations.
  - Long-distance suburban commuters.
- Control stations
  - 16 stations matched on ridership and peak patterns.
  - Not EBD-treated in 2016.



Data

- Coverage: Universe of trip-level smartcard records: card ID, tap-in/tap-out station and time.
- Dates: Five non-consecutive weeks in 2016 (Mar, Jun, Sep, Nov, Dec); 21 working days.
- Time: Morning peak 6:30–8:30 AM, partitioned into eight 15-minute bins.
- Market = OD pair $\times$ day; product = 15-min time bin.
- Sample: $\approx$3.8M trips in 40,191 markets (1,621 EBD + 1,920 control unique OD pairs).

▸ Summary statistics

Measurement

Optimal Route, Train Assignment, and Crowding

- Challenge: We observe only each passenger's **entry/exit station** and **tap-in/tap-out time** — not their route or real-time location.

Optimal Route, Train Assignment, and Crowding

- Challenge: We observe only each passenger's **entry/exit station** and **tap-in/tap-out time** — not their route or real-time location.
- Optimal route: Compute the fastest route for each OD pair using Dijkstra's algorithm on the subway timetable.

Optimal Route, Train Assignment, and Crowding

- Challenge: We observe only each passenger's **entry/exit station** and **tap-in/tap-out time** — not their route or real-time location.
- Optimal route: Compute the fastest route for each OD pair using Dijkstra's algorithm on the subway timetable.
- Travel time decomposition (only TT_{jdt}, ST_{jdt} observed; AT^O, ET^D, RT^R unknown):

$$TT_{jdt} = \underbrace{ST_{jdt}}_{\text{in-train}} + \underbrace{AT^O}_{\text{entry}} + \underbrace{\sum_R RT^R}_{\text{transfers}} + \underbrace{ET^D}_{\text{exit}}$$

Optimal Route, Train Assignment, and Crowding

- Challenge: We observe only each passenger's **entry/exit station** and **tap-in/tap-out time** — not their route or real-time location.
- Optimal route: Compute the fastest route for each OD pair using Dijkstra's algorithm on the subway timetable.
- Travel time decomposition (only TT_{jdt}, ST_{jdt} observed; AT^O, ET^D, RT^R unknown):

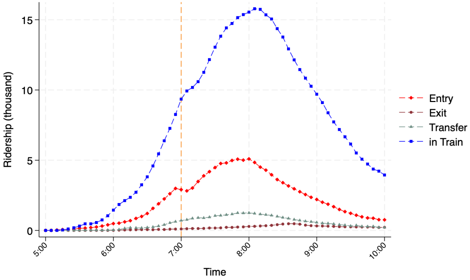
$$TT_{jdt} = \underbrace{ST_{jdt}}_{\text{in-train}} + \underbrace{AT^O}_{\text{entry}} + \underbrace{\textstyle\sum_R RT^R}_{\text{transfers}} + \underbrace{ET^D}_{\text{exit}}$$

- Iterative calibration: Initial guess of AT^O, ET^D, RT^R; predict total travel time for every trip; update parameters to match observed durations; repeat until convergence.

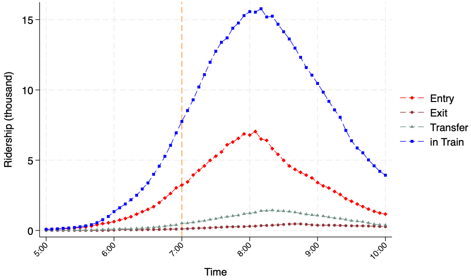
Optimal Route, Train Assignment, and Crowding

- Challenge: We observe only each passenger's **entry/exit station** and **tap-in/tap-out time** — not their route or real-time location.
- Optimal route: Compute the fastest route for each OD pair using Dijkstra's algorithm on the subway timetable.
- Travel time decomposition (only TT_{jdt}, ST_{jdt} observed; AT^O, ET^D, RT^R unknown):
$$TT_{jdt} = \underbrace{ST_{jdt}}_{\text{in-train}} + \underbrace{AT^O}_{\text{entry}} + \underbrace{\textstyle\sum_R RT^R}_{\text{transfers}} + \underbrace{ET^D}_{\text{exit}}$$
- Iterative calibration: Initial guess of AT^O, ET^D, RT^R; predict total travel time for every trip; update parameters to match observed durations; repeat until convergence.
- Real-time location & density: Route + time decomposition ⇒ density per segment × 5-min bin; aggregate to OD-day-15min product. In-station time Route details

EBD Creates Bunching: Departure Time



(a) EBD Stations



(b) Control Stations

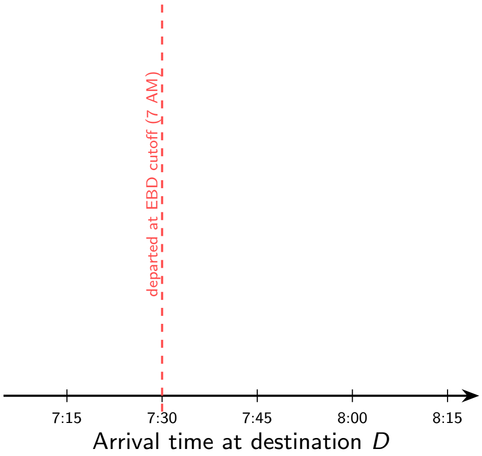
- Total ridership by 5-min tap-in bin; EBD stations vs. control.
- EBD creates visible bunching before the 7 AM discount cutoff.

▸ Bunching in crowding ▸ Crowding by station

Stylized Facts

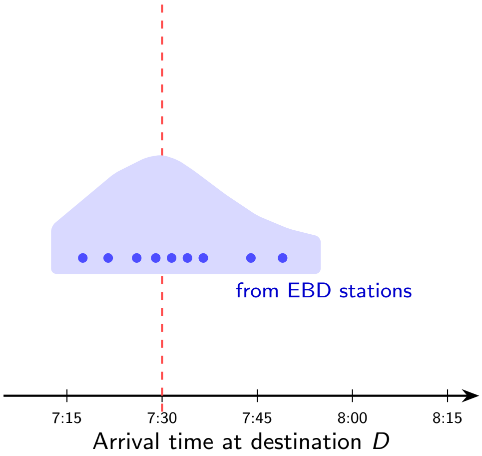
- **Crowding is predictable.** Commuters form rational expectations: $R^2 > 0.98$ on density with OD-time FE. ▸ Evidence
- **Crowding does not slow down the train.** Delay cost < 0.3 RMB even at 90th-pct crowding (VoT $=$ half median wage). ▸ Evidence
- **Crowding does not result in rerouting.** Travel-time distributions are tightly centered around the OD mean, with no fat right tail. ▸ Evidence

Optimal Arrival Time and Rescheduling



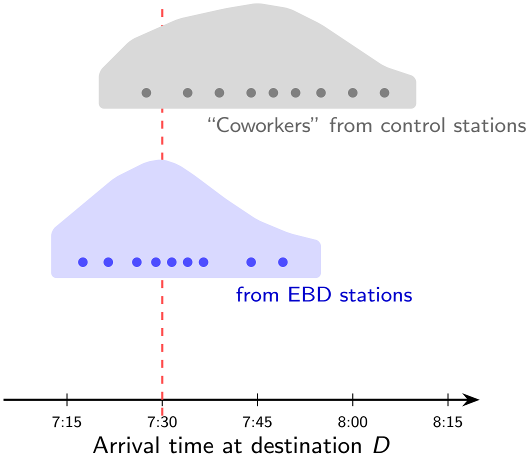
- **Challenge:** Optimal arrival time t^{OA} is unobserved.

Optimal Arrival Time and Rescheduling



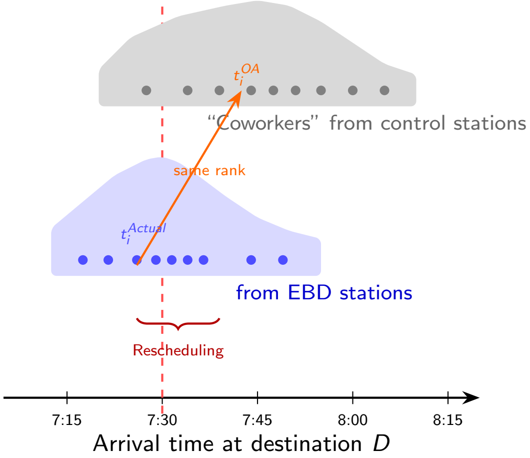
- **Challenge:** Optimal arrival time t^{OA} is unobserved.
- EBD discount incentivizes departure before 7 AM $\Rightarrow$ earlier arrivals at destination.

Optimal Arrival Time and Rescheduling



- **Challenge:** Optimal arrival time t^{OA} is unobserved.
- EBD discount incentivizes departure before 7 AM ⇒ earlier arrivals at destination.
- **Control-as-Proxy:** Control riders (same destination, same day) are unaffected ⇒ their arrival distribution reflects *unconstrained* behavior; EBD riders are *shifted earlier*.

Optimal Arrival Time and Rescheduling

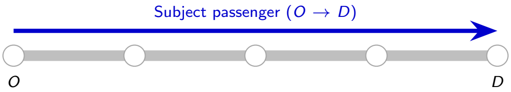


- **Challenge:** Optimal arrival time t^{OA} is unobserved.
- EBD discount incentivizes departure before 7 AM $\Rightarrow$ earlier arrivals at destination.
- **Control-as-Proxy:** Control riders (same destination, same day) are unaffected $\Rightarrow$ their arrival distribution reflects *unconstrained* behavior; EBD riders are *shifted earlier*.
- **Quantile matching $\Rightarrow$ Rescheduling:** Map each EBD rider to the same rank in the control distribution, $t_i^{OA} = F_{\text{Ctrl}}^{-1}(\cdot)$; rescheduling $= \left| t_i^{Actual} - t_i^{OA} \right|$ (zero for controls by construction).

Validation by entry time

Accidental Companion Passengers (ACP)

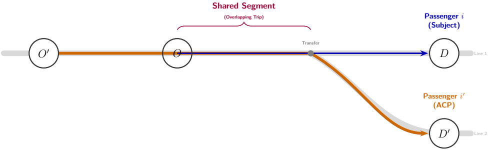
- Source of endogeneity: **Correlated demand shocks**.



Stage 1: subject's trip.

Accidental Companion Passengers (ACP)

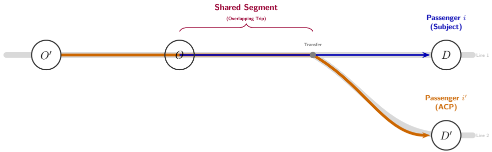
- Source of endogeneity: **Correlated demand shocks**.
- **ACP:** Passengers who share a segment with the focal commuter but travel:
  - From a different origin $O' \neq O$,
  - To a distant destination,
  - travel for different purpose (non-regular commuters).



Stage 2: subject's trip with ACP overlay.

Accidental Companion Passengers (ACP)

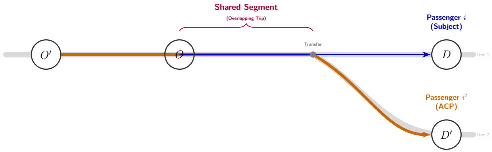
- Source of endogeneity: **Correlated demand shocks**.
- **ACP:** Passengers who share a segment with the focal commuter but travel:
  - From a different origin $O' \neq O$,
  - To a distant destination,
  - travel for different purpose (non-regular commuters).
- **ACT:** Total in-train overlap time with ACPs.



Stage 2: subject's trip with ACP overlay.

Accidental Companion Passengers (ACP)

- Source of endogeneity: **Correlated demand shocks**.
- **ACP:** Passengers who share a segment with the focal commuter but travel:
  - From a different origin $O' \neq O$,
  - To a distant destination,
  - travel for different purpose (non-regular commuters).
- **ACT:** Total in-train overlap time with ACPs.
- ACT from **infrequent users**: under 4 trips per week.



Stage 2: subject's trip with ACP overlay.

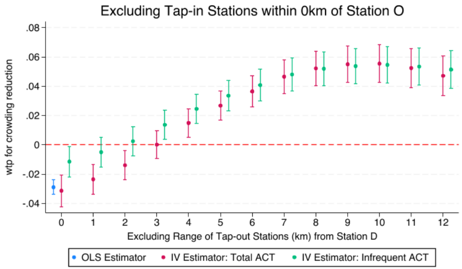
Estimation Results

IV Validity

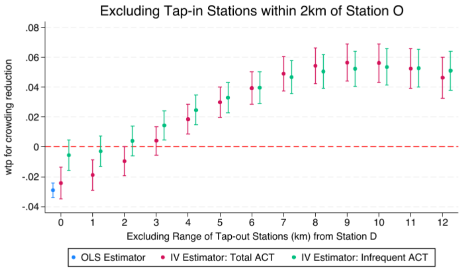
- Source of endogeneity: Correlated demand shocks
  - e.g., everyone needs to arrive at work at the same time.
  - focus on ACPs who are unlikely to share correlated demand shocks with focal passengers.
  - from different O, to different D, with different travel purpose.
  - Turns out that correlated demand shocks are driven by D.

IV Validity

- Source of endogeneity: Correlated demand shocks
  - e.g., everyone needs to arrive at work at the same time.
  - focus on ACPs who are unlikely to share correlated demand shocks with focal passengers.
  - from different O, to different D, with different travel purpose.
  - Turns out that correlated demand shocks are driven by D.



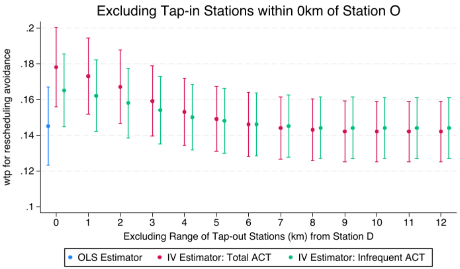
(a) MWTP Crowd, Excl. 0 km from O



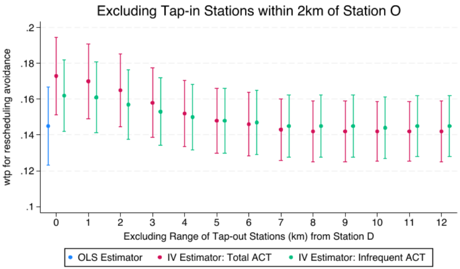
(b) MWTP Crowd, Excl. 2 km from O

IV Validity

- Source of endogeneity: Correlated demand shocks
  - e.g., everyone needs to arrive at work at the same time.
  - focus on ACPs who are unlikely to share correlated demand shocks with focal passengers.
  - from different *O*, to different *D*, with different travel purpose.
  - Turns out that correlated demand shocks are driven by *D*.



(c) MWTP Resch., Excl. 0 km from *O*



(d) MWTP Resch., Excl. 2 km from *O*

Preferred IVs

- Preferred IV
  - ACT from infrequent users
  - with $\|D' - D\| \geq 8$ km.

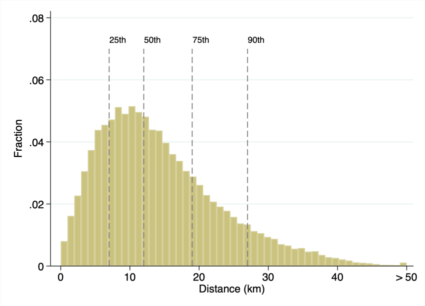
Preferred IVs

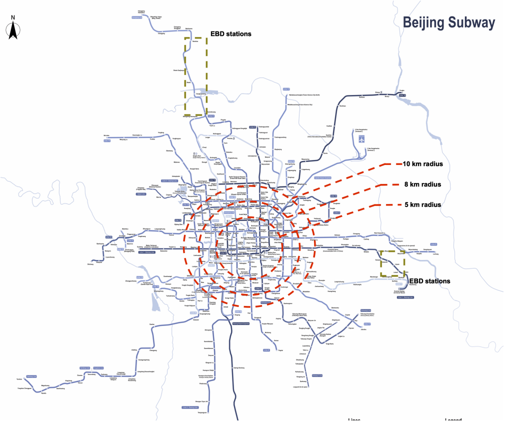
- Preferred IV
  - ACT from infrequent users
  - with $\|D' - D\| \geq 8$ km.

- Additional IVs: EBD-induced timing shifts
  - EBD Stations $\times \mathbf{1}\{t < 7 \text{ AM}\}$
  - EBD Stations $\times$ time trend^{1,2,3}

Preferred IVs

- Preferred IV
  - ACT from infrequent users
  - with $\|D' - D\| \geq 8$ km.
- Additional IVs: EBD-induced timing shifts
  - EBD Stations $\times$ $\mathbf{1}\{t < 7 \text{ AM}\}$
  - EBD Stations $\times$ time trend[1,2,3]





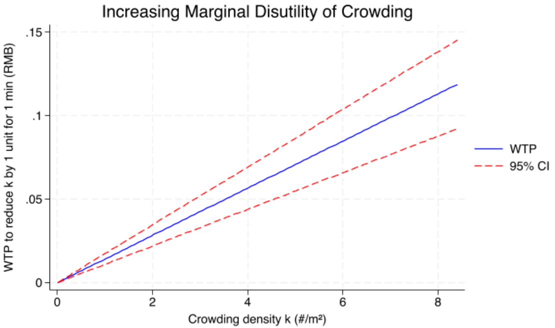
(b) Subway Map: Proportion of Batong Line Riders

GMM Estimation Results (Departure Time Choice by Frequent Riders)

	OLS	GMM: Multi-IV		Nonlinear
		1st Stage	2nd Stage	2nd Stage
	(1)	(2)	(3)	(4)
Crowding	0.0062		−0.0094	
	(0.0003)		(0.0011)	
Crowding^2				−0.000028
				(0.000003)
Price	−0.2093		−0.1821	−0.1513
	(0.0151)		(0.0094)	(0.0090)
Rescheduling	−0.0299	−0.1360	−0.0250	−0.0252
	(0.0010)	(0.0243)	(0.0010)	(0.0010)
ACT		11.1407		
		(0.4645)		
Other IVs		✓		
Wald F-Test			284.8	441.8
R^2	0.6588		−0.0720	
OD-Day FE	X	X	X	X
Time-Day FE	X	X	X	X
Mean Rescheduling	4.654		4.654	
Mean Crowding	127.1		127.1	
Implied MWTP				
MWTP for Crowding	−0.0296		0.0516	see right panel
	(0.00266)		(0.0061)	
MWTP for Rescheduling	0.143		0.137	0.1662
	(0.0114)		(0.0089)	(0.0120)

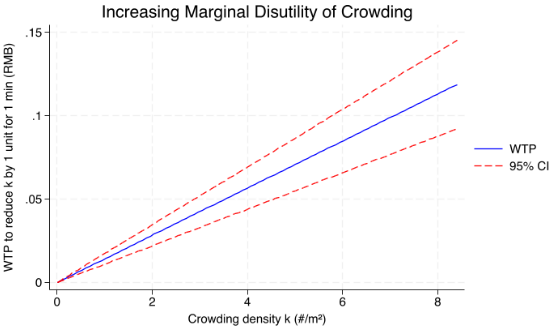
GMM Estimation Results (Departure Time Choice by Frequent Riders)

	OLS	GMM: Multi-IV		Nonlinear
		1st Stage	2nd Stage	2nd Stage
	(1)	(2)	(3)	(4)
Crowding	0.0062		−0.0094	
	(0.0003)		(0.0011)	
$Crowding^2$				−0.000028
				(0.000003)
Price	−0.2093		−0.1821	−0.1513
	(0.0151)		(0.0094)	(0.0090)
Rescheduling	−0.0299	−0.1360	−0.0250	−0.0252
	(0.0010)	(0.0243)	(0.0010)	(0.0010)
ACT		11.1407		
		(0.4645)		
Other IVs		✓		
Wald F-Test			284.8	441.8
R^2	0.6588		−0.0720	
OD-Day FE	X	X	X	X
Time-Day FE	X	X	X	X
Mean Rescheduling	4.654		4.654	
Mean Crowding	127.1		127.1	
Implied MWTP				
MWTP for Crowding	−0.0296		0.0516	see right
	(0.00266)		(0.0061)	panel
MWTP for Rescheduling	0.143		0.137	0.1662
	(0.0114)		(0.0089)	(0.0120)



GMM Estimation Results (Departure Time Choice by Frequent Riders)

	OLS	GMM: Multi-IV		Nonlinear
		1st Stage	2nd Stage	2nd Stage
	(1)	(2)	(3)	(4)
Crowding	0.0062		−0.0094	
	(0.0003)		(0.0011)	
Crowding^2				−0.000028
				(0.000003)
Price	−0.2093		−0.1821	−0.1513
	(0.0151)		(0.0094)	(0.0090)
Rescheduling	−0.0299	−0.1360	−0.0250	−0.0252
	(0.0010)	(0.0243)	(0.0010)	(0.0010)
ACT		11.1407		
		(0.4645)		
Other IVs		✓		
Wald *F*-Test			284.8	441.8
R^2	0.6588		−0.0720	
OD-Day FE	X	X	X	X
Time-Day FE	X	X	X	X
Mean Rescheduling	4.654		4.654	
Mean Crowding	127.1		127.1	
Implied MWTP				
MWTP for Crowding	−0.0296		0.0516	see right
	(0.00266)		(0.0061)	panel
MWTP for Rescheduling	0.143		0.137	0.1662
	(0.0114)		(0.0089)	(0.0120)



- **Crowding:** avg. trip ≈ 32 min ⇒ WTP ≈ 1.7 RMB to reduce crowding by 1 person/m^2. At 8 persons/m^2, MWTP ≈ 3.7 RMB.
- **Rescheduling:** relatively low; 15-min deviation from optimal ≈ 2.1 RMB.
- **Benchmark:** median VoT over the 32-min trip ≈ 11.7 RMB.

▸ Single-IV results ▸ In-station robustness

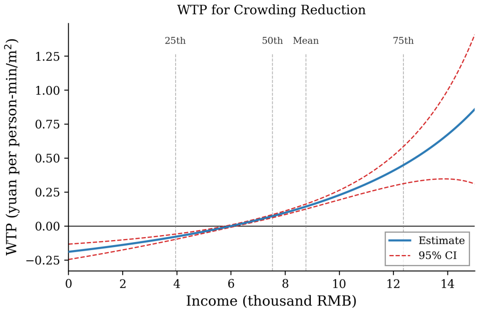
Random Coefficient Model: Structural Parameters

Dep. var.: $\log S_{jdt} - \log S_{jd0}$		
Panel A: Structural Parameters		
Price (α_0)	-0.1787	(0.0109)
Price $\times$ Income (α_1)	$+0.0131$	(0.0048)
Crowding (β_0)	-0.0256	(0.0006)
Crowding $\times$ Income (β_1)	-0.0093	(0.0002)
Rescheduling (ρ_0)	-0.0173	(0.0058)
Rescheduling $\times$ Income (ρ_1)	$+0.0016$	(0.0019)

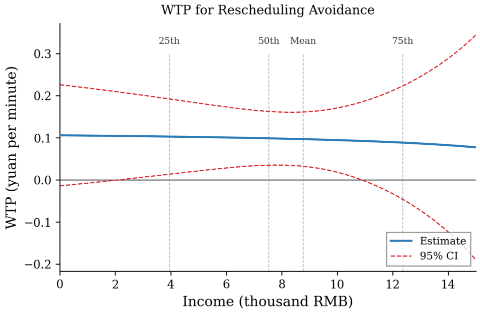
- $\alpha_1 > 0$: higher-income riders less price sensitive.
- $\beta_1 < 0$: higher-income riders have *stronger* crowding disutility.
- ρ_1 not significantly different from zero: rescheduling cost similar across incomes.

Income distribution

WTP by Income



(a) Crowding WTP



(b) Rescheduling WTP

- **Crowding MWTP:** strongly increasing with income. At median income (≈7.5k RMB): ≈ 0.07 RMB/(person·min/m^2). At 90th pct (≈17.5k RMB): ≈ 1.65 RMB/(person·min/m^2).
- **Rescheduling MWTP:** relatively flat (≈ 0.10 RMB/min) across income.

Counterfactuals & Welfare

Welfare Model

- A fixed group of commuters need to travel from O to D.
  - Abstract from the system network and different scheduling.
- Two income groups $g \in \{\ell, h\}$, logit demand:

$$V_g = A + \alpha_g P + \beta_g \text{Crowd},$$

$$\text{Crowd} = \underbrace{\frac{\overline{\text{TT}}}{\text{CarArea} \times J}}_{L:\ \text{crowd exposure}} \cdot N\,; \quad \underbrace{s_g = \frac{e^{V_g}}{1 + e^{V_g}}}_{\text{subway share}}$$

- Consumer surplus: $CS_g = -\frac{M_g}{\alpha_g} \ln(1 + e^{A + \alpha_g P + \beta_g LN})$
- Total welfare: $W = CS + \text{TaxRev} + \Delta\text{MECC}$.

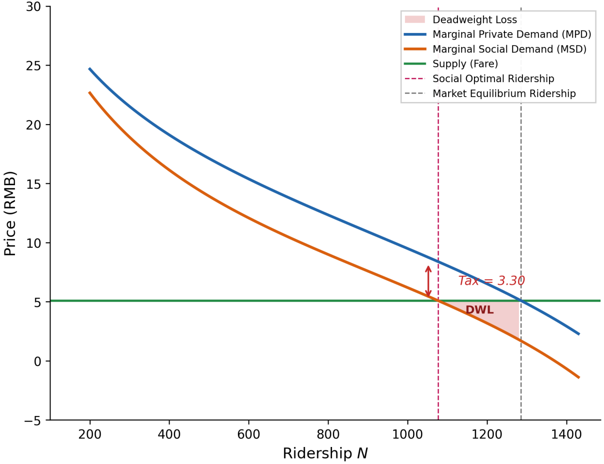
MECC details

Counterfactual Policies

1. **Optimal crowding tax (Pigouvian).**
   - Charge riders the marginal external cost of crowding.
   - Achieves within-system efficiency (not considering MECC) with tax revenue.
2. **Quantity control — queuing.**
   - Fixed boarding capacity; excess demand queues with endogenous wait time.
   - Same ridership reduction as the Pigouvian tax, but no revenue and large DWL.
3. **Two-class configuration.**
   - Business (3 cars) and Standard (3 cars); revenue neutral.
   - Nested logit with nesting parameter λ; induces self-selection by income.

▸ Calibration

Optimal Crowding Tax



- Market equilibrium: $N_{\text{eq}} = 1{,}286$; crowding density $k = 4.03$ persons/m^2.
- Social optimum: $N^{\star} = 1{,}077$; Pigouvian tax $T^{\star} = 3.30$ RMB; ridership $\downarrow$ 16%.
- Externality wedge: $w(N) = L \sum_g (\beta_g/\alpha_g) N_g$.

Welfare Comparisons: 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

Notes: Single-class instruments reduce ridership to $N^\star = 1{,}077$ ($\approx$16% drop). Two-class pricing *increases* total ridership ($\approx$+18%) via Standard-class expansion. Low-income riders dominate Standard; high-income self-select into Business.

Welfare Comparisons: 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

Notes: Single-class instruments reduce ridership to $N^\star = 1{,}077$ ($\approx$16% drop). Two-class pricing *increases* total ridership ($\approx$+18%) via Standard-class expansion. Low-income riders dominate Standard; high-income self-select into Business.

Welfare Comparisons: 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

Notes: Single-class instruments reduce ridership to $N^\star = 1{,}077$ ($\approx$16% drop). Two-class pricing *increases* total ridership ($\approx$+18%) via Standard-class expansion. Low-income riders dominate Standard; high-income self-select into Business.

Welfare Comparisons: 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

Notes: Single-class instruments reduce ridership to $N^\star = 1{,}077$ ($\approx$16% drop). Two-class pricing *increases* total ridership ($\approx$+18%) via Standard-class expansion. Low-income riders dominate Standard; high-income self-select into Business.

Welfare Comparisons: Revenue and Welfare

	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: Pigouvian tax most efficient single-class instrument (revenue redistributable). Queuing dissipates the would-be tax revenue as DWL (≈3,000 RMB), and is regressive. Two-class pricing yields the highest welfare; *both* groups gain ($\Delta W_\ell = +272$, $\Delta W_h = +1{,}399$, in 1,000 RMB). Single-class instruments are regressive.

Welfare Comparisons: Revenue and Welfare

	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: Pigouvian tax most efficient single-class instrument (revenue redistributable). Queuing dissipates the would-be tax revenue as DWL ($\approx$3,000 RMB), and is regressive. Two-class pricing yields the highest welfare; *both* groups gain ($\Delta W_\ell = +272$, $\Delta W_h = +1{,}399$, in 1,000 RMB). Single-class instruments are regressive.

Welfare Comparisons: Revenue and Welfare

	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: Pigouvian tax most efficient single-class instrument (revenue redistributable). Queuing dissipates the would-be tax revenue as DWL (≈3,000 RMB), and is regressive. Two-class pricing yields the highest welfare; *both* groups gain ($\Delta W_\ell = +272$, $\Delta W_h = +1{,}399$, in 1,000 RMB). Single-class instruments are regressive.

Welfare Comparisons: Revenue and Welfare

	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: Pigouvian tax most efficient single-class instrument (revenue redistributable). Queuing dissipates the would-be tax revenue as DWL (≈3,000 RMB), and is regressive. Two-class pricing yields the highest welfare; *both* groups gain ($\Delta W_\ell = +272$, $\Delta W_h = +1{,}399$, in 1,000 RMB). Single-class instruments are regressive.

Conclusions

Conclusions

- First causal estimate of monetary WTP to avoid subway crowding.
  - MWTP $= 0.05$ RMB per person/m^2 per minute (at mean crowding and income).
  - MWTP increases with crowding level (convex disutility) and rider income.
  - Total crowding externality ≈ 6 RMB on a representative 32-min trip.
- Policy implications:
  - Optimal crowding tax $T^{\star} \approx 3.30$ RMB; optimal ridership $\approx 16\%$ below baseline.
  - Quantity control (queuing) performs much worse.
  - Two-class configuration with second-degree price discrimination: induces self-selection; raises welfare for both income groups.
- Methodological contributions: Instrument (ACP/ACT) and measurements (train assignment, optimal arrival time) leveraging the network feature of the transit system.

Thank You!

kaiwu@utexas.edu

Appendix

Appendix: Summary Statistics

	EBD Stations		Control Stations	
	Mean	SD	Mean	SD
# passengers (subway)	10.96	21.81	12.41	16.28
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 (person·min/m^2)	121.74	59.36	123.35	45.25
Crowding per min (person/m^2)	3.88	1.35	4.14	1.34
ACP (infrequent)	0.07	0.05	0.07	0.07
ACP (total)	0.70	0.59	0.84	0.85
ACT (infrequent, min)	1.23	0.95	1.09	1.02
ACT (total, min)	11.55	9.85	12.42	12.16
Observations	142,216		179,312	

- Each observation is an OD-day-15-min-bin. 21 working days in 2016.

Back

Appendix: 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	4.43	1.42	2.61	6.23	34,782,743

- Mean exit time normalized to 2 minutes to break collinearity (see iterative calibration).

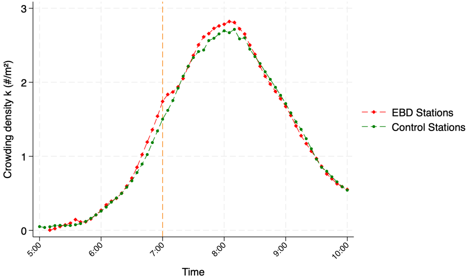
Back

Appendix: Real-time location & crowding density (details)

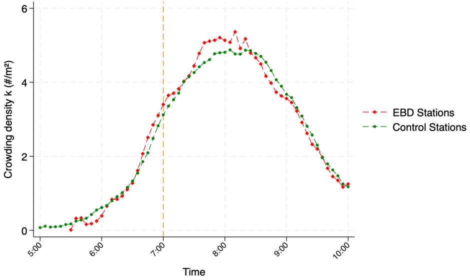
- Real-time location: With the route and time decomposition, we can determine every passenger's real-time location during their trip.
- Discretize into segment × 5-min bins.
  - A trip segment: track between two adjacent stations.
  - Each passenger's presence is assigned to bins, weighted by time spent.
- Crowding density: Passengers per segment-time bin, divided by total train floor area (from train specs and schedule).
- Product-level aggregation: Average crowding within each OD-day-15min (departure time) bin.
- Expected crowding: Predicted from OD-DOW-time-bin FE ($R^2 > 0.98$).

Back

Appendix: EBD Bunching — Crowding



(a) Average Across Stations

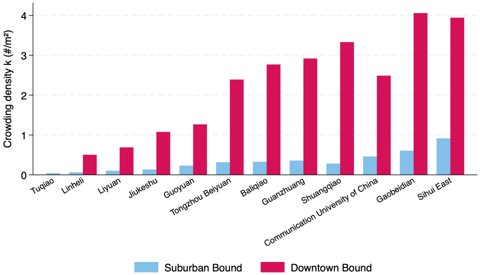


(b) Most Downstream Stations

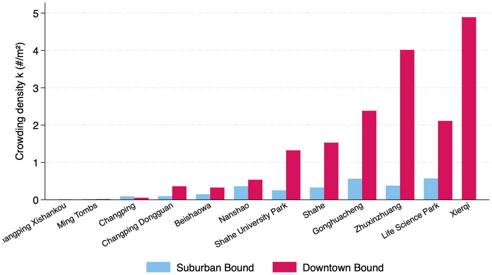
- Downtown-bound in-train crowding (persons/m^2) passing each station over time.
- Crowding is severe and systematic during morning peak hours.

Back

Appendix: Crowding by Station (EBD Lines)



(a) Batong Line



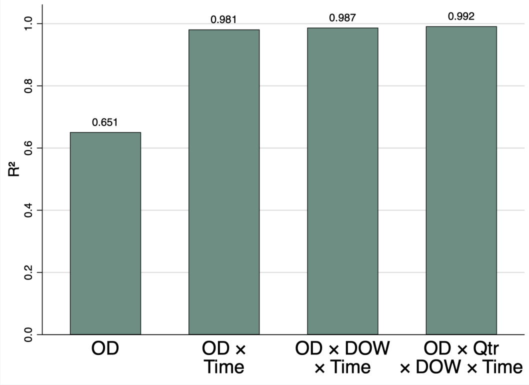
(b) Changping Line

- In-train crowding density (persons/m^2) by station and direction for EBD lines.
- Crowding is severe in the *downtown-bound* direction during morning peak.

Back

Appendix: Crowding is Predictable

– Our model is essentially a full-information model: commuters form rational expectations about crowding.
– This requires that crowding patterns are predictable.

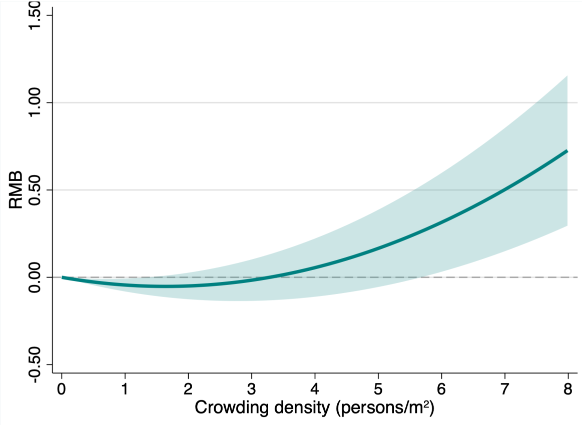


- Dep. var: crowding density (persons/m^2). $R^2 > 0.98$ with OD-time FE.

Back

Appendix: Crowding Does Not Slow Down the Train

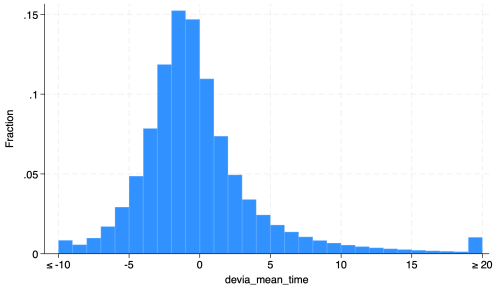
Total monetary cost of crowding due to delay



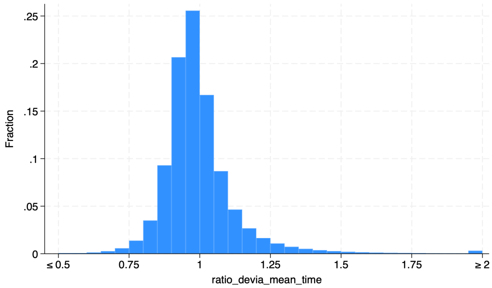
- VoT assumed to be half of median earning.
- Delay cost $<$ 0.3 RMB under the 90th percentile of crowding.
- A tiny fraction of the WTP for crowding reduction.

Back

Appendix: Crowding Does Not Result in Rerouting



(a) Deviation from mean (minutes)

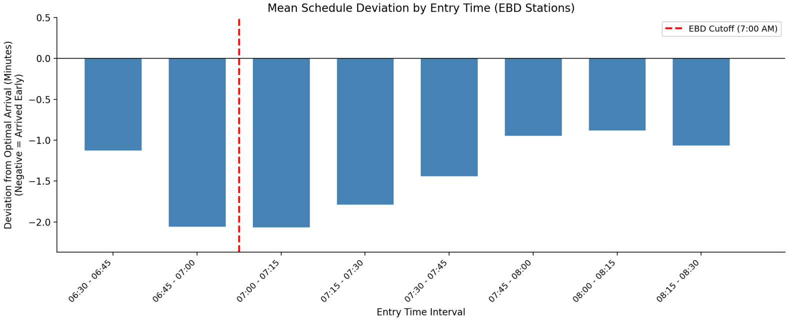


(b) Ratio to mean travel time

- 65.1% of trips within $\pm$3 min of OD mean; 95.8% within $\pm$10 min.
- Rerouting would significantly increase travel time even via the next-best route, but we do not see this in the distribution.

◀ Back

Appendix: Validation — Mean Schedule Deviation by Entry Time



- Before the 7 AM EBD cutoff: riders arrive *earlier* than optimal (≈ -1 to -2 min).
- After 7 AM: deviations smaller — consistent with non-responsive riders.
- Pattern confirms that EBD induces rescheduling behavior.

◂ Back

Appendix: GMM Single-IV Results

	GMM: Single IV (ACT)	
	1st Stage	2nd Stage
	(1)	(2)
Crowding		−0.0201
		(0.0019)
Price	−5.1519	−0.3280
	(0.4331)	(0.0196)
Rescheduling	0.0354	−0.0274
	(0.0306)	(0.0014)
ACT	11.5338	
	(0.4788)	
Wald *F*-Test		580.2
R^2		−0.3272
OD-Day FE	X	X
Time-Day FE	X	X
Mean Rescheduling		4.654
Mean Crowding		127.1
Implied MWTP		
MWTP for Crowding		0.0613
		(0.0055)
MWTP for Rescheduling		0.0856
		(0.0072)

Back

Appendix: Robustness — Controlling for In-Station Crowding

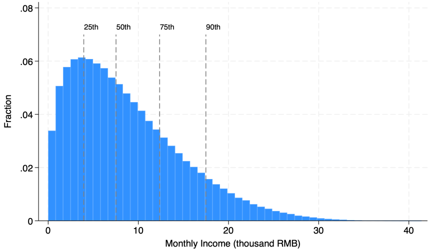
	(1) Multi-IV Baseline	(2) + In-Station Control
In-train crowding	−0.0094	−0.0078
	(0.0011)	(0.0013)
Price	−0.1821	−0.1618
	(0.0094)	(0.0097)
Rescheduling	−0.0250	−0.0242
	(0.0010)	(0.0011)
In-station crowding		−0.0077
		(0.0023)
N	295,400	295,400
Wald *F*	284.82	470.58
MWTP crowding	0.0516	0.0482
	(0.0061)	(0.0075)

- Cols (3)–(4) from Appendix Table B.2. Multi-IV (ACT + EBD policy); col (2) treats in-station crowding as a second *endogenous* regressor.

Back

Appendix: Income Distribution

- **Source:** Baidu commuting data (≈7.6 million commuters, ≈350k grids, imputed income bins).
- **Income distribution for each OD pair:** Map grids to stations via inverse-distance weighting (3 km catchment). Fit beta distributions to grouped income data.
- **For BLP estimation:** 150 simulated income draws per OD-date market.



Median ≈ 7.5k RMB; mean ≈ 8.8k RMB.

Back

Appendix: Marginal External Road Congestion Cost (MECC)

- Riders diverted from subway substitute partly into road modes, increasing road congestion.
- Mode substitution:
  - Low-income: $p_{\text{bus}} = 0.5$, $p_{\text{car}} = 0.25$.
  - High-income: $p_{\text{bus}} = 0.15$, $p_{\text{car}} = 0.75$.
- Congestion factor: $k_g = p_{\text{car},g} + 0.4\, p_{\text{bus},g}$.
  - 1 bus ride $\approx$ 0.4 car trips.
- MECC: $\Delta\text{MECC}_g = c_{\text{car}} \cdot k_g \cdot \bar{d} \cdot (N_g^{\text{base}} - N_g^{\text{cf}})$.
  - c_{car}: congestion cost per passenger-km by automobile (RMB).
  - $\bar{d}$: mean trip distance (km).

Back

Appendix: Calibration

Parameter	Value	Description	Source
Income groups (p41/p59 split)			
y_ℓ	6.19k RMB	Low-income (41st pctile)	Baidu data
y_h	9.11k RMB	High-income (59th pctile)	Baidu data
$\bar{y}$	8.77k RMB	Mean simulated income	Baidu data
$\alpha_\ell,\ \alpha_h$	$-0.213,\ -0.174$	Price sensitivity by group	Computed (BLP)
$\beta_\ell,\ \beta_h$	$-0.002,\ -0.029$	Crowding disutility by group	Computed (BLP)
Market and calibration (ridership-weighted)			
k	4.03 persons/m^2	Average in-train density	Sample mean
$\bar{t}$	31.83 min	Mean in-train travel time	Sample mean
N_{eq}	1,286	Baseline equilibrium ridership	$4.03 \times 53.2 \times 6$
M	2,983	Total market size	$N_{\text{eq}}/0.43$
L	0.100	Crowd exposure	$31.83/(53.2 \times 6)$
P_{base}	5.08 RMB	Baseline fare	Sample mean
A	1.77	Calibrated mean utility	Match N_{eq}
Two-class configuration			
$J_B,\ J_S$	3, 3	Business / Standard cars	Symmetric split
λ	0.5	Nesting parameter	Assumed; sensitivity in appendix

Appendix: BLP Structural Parameters

Parameter	Value	Description	Source
Structural parameters			
α_0	-0.1787	Price coeff. (intercept)	BLP estimation
α_1	$+0.0131$	Price coeff. (income slope)	BLP estimation
β_0	-0.0256	Crowding coeff. (intercept)	BLP estimation
β_1	-0.0093	Crowding coeff. (income slope)	BLP estimation
ρ_0	-0.0173	Rescheduling coeff. (intercept)	BLP estimation
ρ_1	$+0.0016$	Rescheduling coeff. (income slope)	BLP estimation

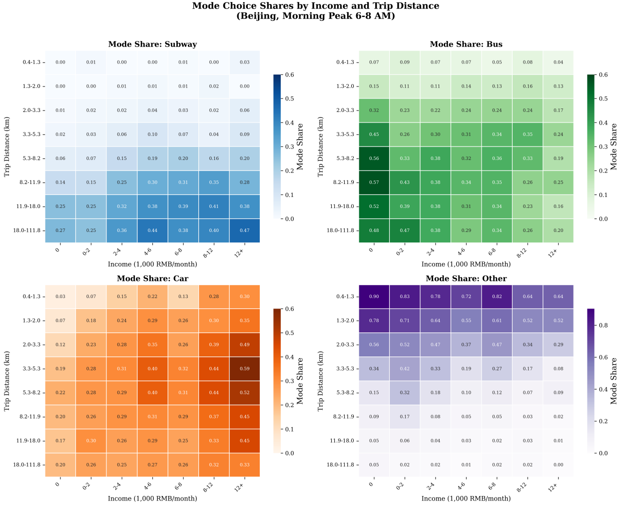
Back

Appendix: Transportation Mode Choice

Multinomial Logit

	Subway	Bus	Car
Income	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)
Distance	0.473	0.437	0.412
	(0.004)	(0.004)	(0.004)
N		80,259	
Pseudo R^2		0.258	

Base: "other" mode. SE in parens. Microdata from Beijing Household Transportation Survey 2015.



Mode shares by income and trip distance.