Enhancing the gravity model for commuters: Time-and-spatial-structure-based improvements in Japan's metropolitan areas

Supplementary Material

Yixuan Y Zheng[1], Hideki Takayasu[2], Misako Takayasu[1,2]*

[1] Department of Systems and Control Engineering, School of Engineering Institute of Sience Tokyo, Yokohama 226-8502, Japan

[2] Department of Computer Science, School of Computing Institute of Sience Tokyo, Yokohama 226-8502, Japan

Contents

1 Transportation mode identification results

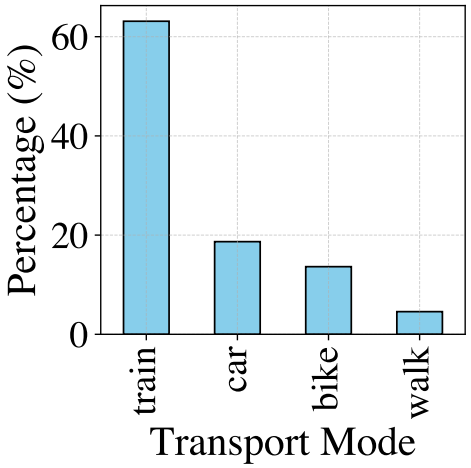


Figure S1: **Distribution of transportation modes identified from smartphone GPS data.** Our dataset shows railway users comprise 60% of trips, compared to 50% in national statistics [4]. Car (or bus, motorcycle) users: 19% (GPS) vs. 14% (national); bicycle users: 13% (GPS) vs. 13% (national); walking: 5% (GPS) vs. 24% (national). The higher proportion of railway users and lower commuters commute by walk in our data likely reflects sampling bias in smartphone data, which potentially underrepresents individuals under 13 and over 65 years of age who may rely more heavily on other transportation modes.

2 Multicollinearity Diagnostics

Table 1 presents the variance inflation factor (VIF) analysis for the selected model specification. All VIF values are below 3, indicating no severe multicollinearity concerns. The highest VIF of 2.84 for $\log(\text{CAP}_2)$ remains well below the conventional threshold of 5-10 that would indicate problematic multicollinearity.

Table 1: Variance inflation factor (VIF) analysis for the selected model

Variable	VIF	Interpretation
NT (squ_trans)	2.50	No concern
CT (gowork_time)	2.72	No concern
log(CD)	2.48	No concern
PK (peak)	1.05	No concern
$\log(\text{CAP}_1)$	1.41	No concern
$\log(\text{CAP}_2)$	2.84	No concern

Note: VIF < 5 indicates no severe multicollinearity [3]

3 Variable Selection Process

To arrive at a robust model specification, we systematically evaluated ten candidate models addressing three key methodological considerations:

(1) Functional form of attributes: We tested both linear and logarithmic transformations of distance and capacity variables to capture potential non-linear effects. Logarithmic specifications can reflect behavioral phenomena such as diminishing marginal disutility—where the perceived cost of an additional kilometer decreases with trip length.

(2) Multicollinearity mitigation: Given the inherent correlation between travel time and distance in transportation networks, we evaluated various variable combinations using variance inflation factor (VIF) diagnostics. High multicollinearity can lead to unstable parameter estimates and unreliable inference. We tested specifications including: time with log-transformed distance, log-transformed time alone,

and distance alone, to identify combinations that maintain model stability while preserving explanatory power.

(3) Transfer station heterogeneity: We compared specifications using a single aggregated capacity variable ($\log(\text{CAP}_1 + \text{CAP}_2)$) against models with separate variables for first and second transfer stations. This comparison tests whether commuters perceive transfers at different points in their journey differently—a critical consideration given Tokyo's hierarchical railway network structure where second transfers often occur at major central hubs.

Model performance was evaluated using McFadden's pseudo-R^2 for explanatory power, and the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for balancing fit with parsimony. Table 2 presents the comparison results.

Table 2: Systematic comparison of alternative model specifications

Parameters	Model specification and rationale	R^2	AIC	BIC
NT, CT	Baseline model with transfer count and travel time	0.068	2.5949E+05	2.949E+05
NT, log(CT)	Tests non-linear time effect; marginal improvement suggests limited non-linearity in time perception	0.078	2.782E+05	2.783E+05
NT, CD	Distance-only specification; inferior performance indicates time dominates distance in route choice	0.053	2.939E+05	2.939E+05
NT, log(CD)	Non-linear distance specification; improved fit suggests decreasing marginal disutility of distance	0.067	2.783E+05	2.783E+05
NT, CT, log(CD)	Combined time-distance model; performance comparable to log(CT) model while avoiding multicollinearity through mixed specification	0.079	2.752E+05	2.752E+05
NT, CT, log(CD), PK	Incorporates peak-hour effects; substantial improvement indicates temporal constraints significantly influence route choice	0.156	2.729E+05	2.729E+05
NT, CT, log(CD), PK, CAP_1, CAP_2	Tests linear capacity effects at transfer stations; limited improvement suggests need for transformation	0.21	2.873E+05	2.873E+05
NT, CT, log(CD), PK, $\log(\text{CAP}_{\text{total}})$	Aggregated log-capacity specification; improved fit confirms non-linear capacity perception	0.161	2.654E+05	2.654E+05
NT, CT, log(CD), PK, $\log(\text{CAP}_1)$, $\log(\text{CAP}_2)$	Disaggregated log-capacity model; superior performance indicates heterogeneous effects of first versus second transfer station capacity	0.232	2.611E+05	2.611E+05

The final parameter set we chose is as follows:

$$\mathbf{X}_{nj} = (\text{CT}_{nj}, \log(\text{CD}_{nj}), \text{PK}_{nj}, \text{NT}_{nj}, \log(\text{CAP}_{nj1}), \log(\text{CAP}_{nj2})) \tag{1}$$

This specification not only resolves the multicollinearity issue (reducing VIF values to acceptable levels) but also has strong behavioral foundations. The logarithmic distance transformation captures decreasing marginal disutility—an additional kilometer matters more for short trips than long ones. Similarly, logarithmic capacity reflects diminishing returns to station size, where improvements at congested stations provide greater benefit than expansions at already-large facilities.

4 CPOD results for the rest 12 months

Table 3: Jan: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.480	0.130	246.011
No. of Transfers (NT_{nj})	0.591^{***}	0.017	135.431
Commuting Time (CT_{nj}, min)	0.065^{***}	0.027	48.006
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.613^{***}	0.038	127.530
Peak Departure (PK_{nj}, dummy)	-0.114^{***}	0.006	-175.050
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.548^{*}	0.020	2.206
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.350^{***}	0.015	-179.897

Number of users	91,285
Number of routes	16,665
Number of ODs	4226
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-134,733.081$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-103,585.409$
AIC	207,184.818
McFadden's ρ^2	0.231
Hit rate	55.8%
Spearman ρ	0.724

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 4: Feb: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.594	0.119	274.275
No. of Transfers (NT_{nj})	0.591^{***}	0.015	153.400
Commuting Time (CT_{nj}, min)	0.102^{***}	0.021	97.812
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.484^{***}	0.041	94.400
Peak Departure (PK_{nj}, dummy)	-0.134^{***}	0.006	-203.795
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.524^{*}	0.018	2.426
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.360^{***}	0.014	-206.615

Number of users	96,988
Number of routes	18,310
Number of ODs	4548
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-140,258.572$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-107,633.099$
AIC	215,280.199
McFadden's ρ^2	0.233
Hit rate	55.8%
Spearman ρ	0.724

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5: Mar: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.721	0.146	231.007
No. of Transfers (NT_{nj})	0.552^{***}	0.016	124.112
Commuting Time (CT_{nj}, min)	0.103^{***}	0.024	85.284
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.572^{***}	0.043	106.764
Peak Departure (PK_{nj}, dummy)	-0.110^{***}	0.006	-164.828
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.479^{*}	0.019	2.035
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.319^{***}	0.015	-170.726

Number of users	124,398
Number of routes	22,887
Number of ODs	5608
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-109,871.924$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-86,603.967$
AIC	173,221.935
McFadden's ρ^2	0.212
Hit rate	54.8%
Spearman ρ	0.721

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 6: April: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.658	0.116	285.841
No. of Transfers (NT_{nj})	0.569^{***}	0.014	156.309
Commuting Time (CT_{nj}, min)	0.087^{***}	0.022	80.378
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.579^{***}	0.034	133.157
Peak Departure (PK_{nj}, dummy)	-0.117^{***}	0.005	-209.778
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.523^{*}	0.017	2.543
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.340^{***}	0.013	-210.615

Number of users	125,368
Number of routes	23,887
Number of ODs	5908
Initial log-likelihood $\mathcal{L}(\hat{\beta})$	$-170,016.797$
Final log-likelihood $\mathcal{L}(\beta_0)$	$-130,530.778$
AIC	261,075.555
McFadden's ρ^2	0.232
Hit rate	55.8%
Spearman ρ	0.724

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 7: May: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.746	0.144	235.386
No. of Transfers (NT_{nj})	0.582^{***}	0.016	141.022
Commuting Time (CT_{nj}, min)	0.108^{***}	0.022	97.443
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.490^{***}	0.048	82.222
Peak Departure (PK_{nj}, dummy)	-0.127^{***}	0.006	-180.138
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.499^{*}	0.019	2.105
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.349^{***}	0.015	-188.242

Number of users	86,303
Number of routes	15,791
Number of ODs	3980
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-116,432.050$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-90,225.163$
AIC	180,464.327
McFadden's ρ^2	0.225
Hit rate	54.9%
Spearman ρ	0.722

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 8: Jun: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.656	0.156	213.444
No. of Transfers (NT_{nj})	0.557^{***}	0.018	113.763
Commuting Time (CT_{nj}, min)	0.133^{***}	0.019	138.972
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.336^{***}	0.052	51.596
Peak Departure (PK_{nj}, dummy)	-0.126^{***}	0.006	-165.100
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.504^{*}	0.021	1.970
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.333^{***}	0.017	-159.781

Number of users	103,379
Number of routes	19,260
Number of ODs	4769
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-99,185.420$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-77,155.670$
AIC	154,325.340
McFadden's ρ^2	0.222
Hit rate	55.2%
Spearman ρ	0.720

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 9: Jul: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.618	0.147	223.666
No. of Transfers (NT_{nj})	0.558^{***}	0.018	119.221
Commuting Time (CT_{nj}, min)	0.094^{***}	0.026	73.407
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.594^{***}	0.043	109.709
Peak Departure (PK_{nj}, dummy)	-0.102^{***}	0.006	-151.900
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.492^{*}	0.020	1.964
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.327^{***}	0.016	-162.602

Number of users	81,681
Number of routes	15,333
Number of ODs	3810
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-107,459.864$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-84,659.186$
AIC	169,332.372
McFadden's ρ^2	0.212
Hit rate	54.8%
Spearman ρ	0.723

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 10: Aug: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.461	0.141	224.660
No. of Transfers (NT_{nj})	0.596^{***}	0.019	125.285
Commuting Time (CT_{nj}, min)	0.090^{***}	0.028	64.458
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.534^{***}	0.049	87.657
Peak Departure (PK_{nj}, dummy)	-0.119^{***}	0.006	-162.589
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.533^{*}	0.022	2.00
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.356^{***}	0.017	-168.810

Number of users	69,379
Number of routes	13,167
Number of ODs	3268
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-110,494.638$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-86,174.288$
AIC	172,362.575
McFadden's ρ^2	0.220
Hit rate	55.2%
Spearman ρ	0.722

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 11: Sep: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.797	0.138	249.497
No. of Transfers (NT_{nj})	0.562^{***}	0.016	130.386
Commuting Time (CT_{nj}, min)	0.103^{***}	0.022	92.981
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.542^{***}	0.042	102.724
Peak Departure (PK_{nj}, dummy)	-0.108^{***}	0.005	-179.773
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.504^{*}	0.018	2.196
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.326^{***}	0.014	-180.995

Number of users	72,981
Number of routes	14,130
Number of ODs	3488
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-115,982.037$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-88,833.148$
AIC	177,680.295
McFadden's ρ^2	0.234
Hit rate	55.9%
Spearman ρ	0.726

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 12: Oct: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.189	0.123	239.810
No. of Transfers (NT_{nj})	0.626^{***}	0.021	126.139
Commuting Time (CT_{nj}, min)	0.042^{***}	0.035	24.291
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.570^{***}	0.045	102.033
Peak Departure (PK_{nj}, dummy)	-0.113^{***}	0.006	-154.291
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.604^{*}	0.022	2.187
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.385^{***}	0.018	-175.552

Number of users	86,307
Number of routes	16,327
Number of ODs	4070
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-122,187.512$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-93,990.382$
AIC	187,994.764
McFadden's ρ^2	0.231
Hit rate	55.9%
Spearman ρ	0.724

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 13: Nov: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.505	0.132	242.723
No. of Transfers (NT_{nj})	0.573^{***}	0.017	130.445
Commuting Time (CT_{nj}, min)	0.086^{***}	0.024	70.713
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.552^{***}	0.043	102.492
Peak Departure (PK_{nj}, dummy)	-0.120^{***}	0.006	-179.158
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.551^{*}	0.020	2.185
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.342^{***}	0.015	-177.479

Number of users	81,226
Number of routes	15,492
Number of ODs	3582
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-132,129.742$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-101,341.586$
AIC	202,697.172
McFadden's ρ^2	0.233
Hit rate	55.9%
Spearman ρ	0.725

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 14: Dec: Estimation Results of Railway Commuting Path Choice Model

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.720	0.165	205.237
No. of Transfers (NT_{nj})	0.563^{***}	0.019	111.219
Commuting Time (CT_{nj}, min)	0.113^{***}	0.025	89.374
Commuting Distance ($\log(\mathrm{CD}_{\mathrm{nj}})$, km)	0.494^{***}	0.051	77.767
Peak Departure (PK_{nj}, dummy)	-0.111^{***}	0.006	-150.784
Capacity at 1st transfer ($\log(\mathrm{CAP}_{\mathrm{nj1}})$)	0.503^{*}	0.022	1.830
Capacity at 2nd transfer ($\log(\mathrm{CAP}_{\mathrm{nj2}})$)	-0.325^{***}	0.017	-153.058

Number of users	79,180
Number of routes	14,931
Number of ODs	3719
Initial log-likelihood $\mathcal{L}(\beta_0)$	$-93,832.281$
Final log-likelihood $\mathcal{L}(\hat{\beta})$	$-72,452.673$
AIC	144,919.346
McFadden's ρ^2	0.228
Hit rate	55.3%
Spearman ρ	0.723

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

5 Alternative Route choice maps

5.1 High-capacity transfer station aversion

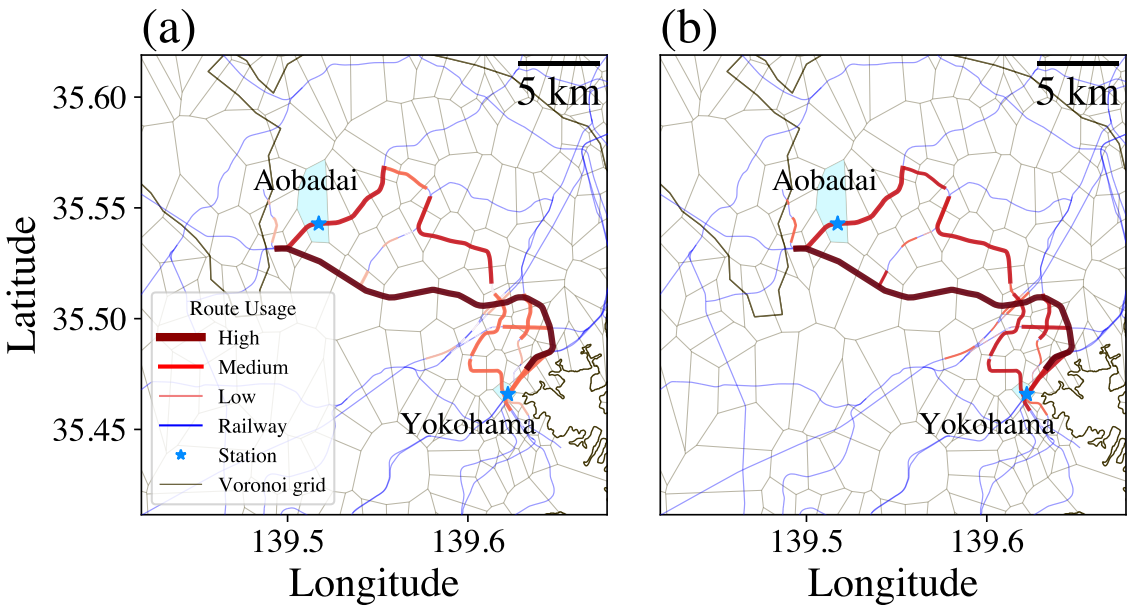


Figure S2: **Real and model-predicted route choice maps for the Aobadai to Yokohama OD pair.** (a) Actual route usage based on GPS data showing high, medium, and low-usage routes. (b) Route usage predicted by the canonical model. Line thickness and darkness indicate usage intensity. The rarker red route is R1, the lighter red route is R2. Two primary route alternatives are examined (detailed in Table 15). The model successfully captures the most and least utilized routes, with predicted patterns closely resembling observed usage. **Notably, commuters avoid crowded transfer stations (R1's transfer station) and prefer R2, which offers better seating availability, despite similar travel times.**

Table 15: Route options from left to right visualized in Fig. 2 for Aobadai to Yokohama OD pair

Route ID	Line Sequence	Travel Time (min)	Transfers
R1	Den-en-toshi Line → Yokohama Line	46	1
R2	Den-en-toshi Line → Yokohama Municipal Subway Blue Line	47	1

Note: Transfer stations and their approximate daily passenger capacity rankings are as follows: (1) Nagatsuda Station (1st transfer for R1); (2) Azamino Station (1st transfer for R2).

5.2 Transfer station aversion

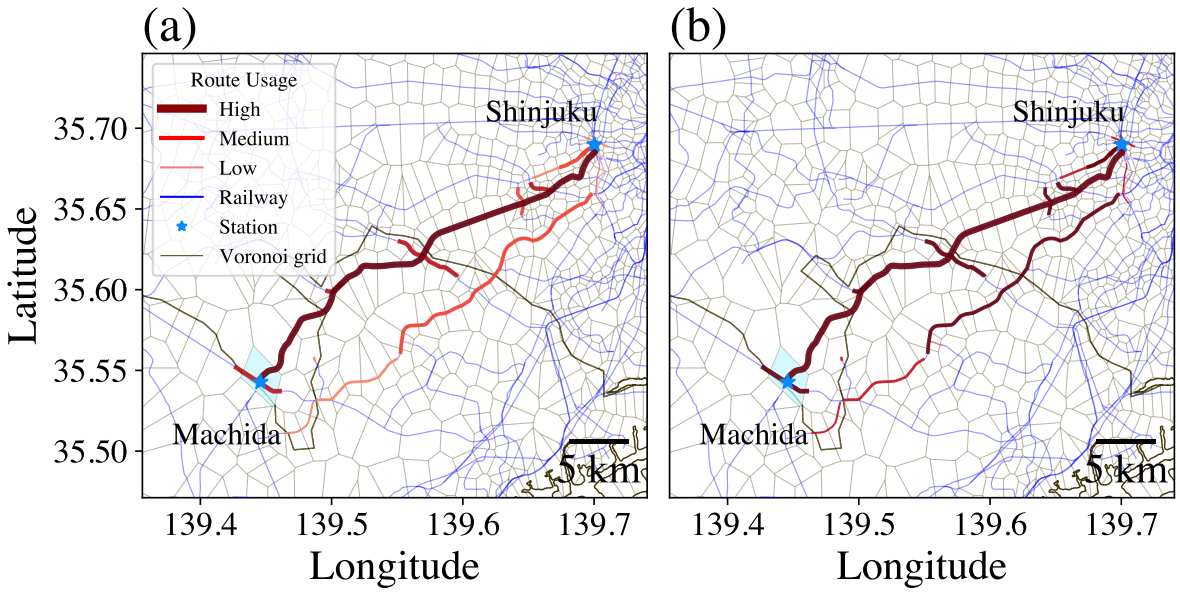


Figure S3: **Real and model-predicted route choice maps for the Machida to Shinjuku OD pair.** (a) Actual route usage based on GPS data showing high, medium, and low-usage routes. (b) Route usage predicted by the canonical model. Line thickness and darkness indicate usage intensity. The rarker red route is R1, the lighter red route is R2. Two primary route alternatives are examined (detailed in Table 16). The model successfully captures the most and least utilized routes, with predicted patterns closely resembling observed usage. **Notably, commuters avoiding the most crowded route prefer alternatives with two transfers, likely because each transfer provides additional opportunities for seating.**

Table 16: Route options from left to right visualized in Fig. 2 for Machida to Shinjuku OD pair

Route ID	Line Sequence	Travel Time (min)	Transfers
R1	Odakyu Line	55	0
R2	Yokohama Line → Den-en-toshi Line → Yamanote Line	67	2

Note: Transfer stations and their approximate daily passenger capacity rankings are as follows: (1) Shinjuku Station (2st transfer for R2); (2) Nagatsuda Station (1st transfer for R2).

6 Data bias

We acknowledge that our results depend on the quality of the GPS dataset. To assess potential bias, we validated our population estimates against official statistics from the Ministry of Land, Infrastructure, Transport and Tourism of Japan [4]. As shown in Fig. S4, our monthly (around 30 days) GPS-derived population estimates are approximately half of the official population counts, with the relationship being consistently proportional across different areas. This consistent ratio suggests no obvious spatial bias in our GPS sample, as the underrepresentation appears uniform rather than concentrated in specific demographic or geographic regions.

7 Commuter Trajectory

To demonstrate the reliability of "home" and "work" status' identification, we present a typical commuter's trajectory, as depicted below (Figure S5):

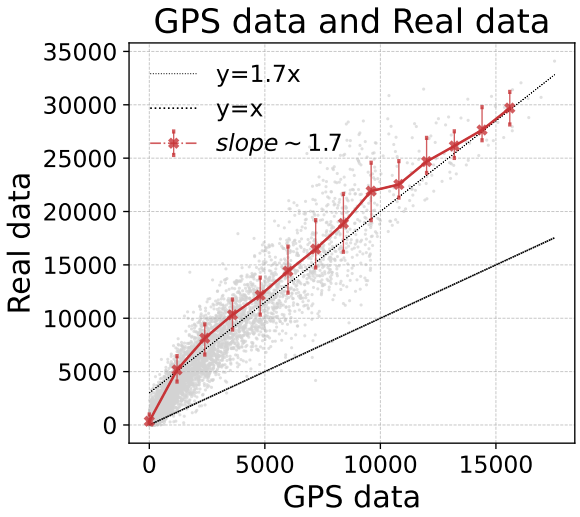
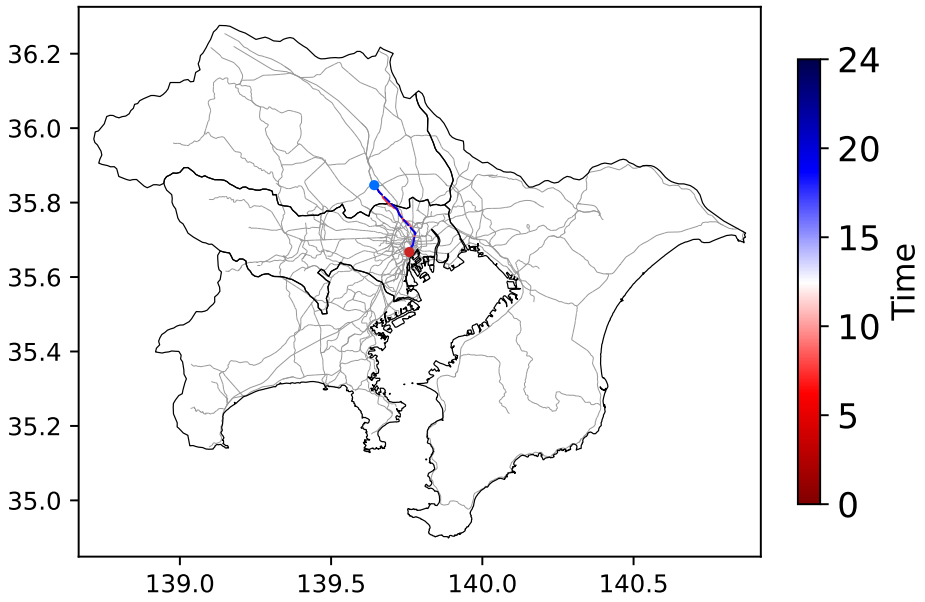
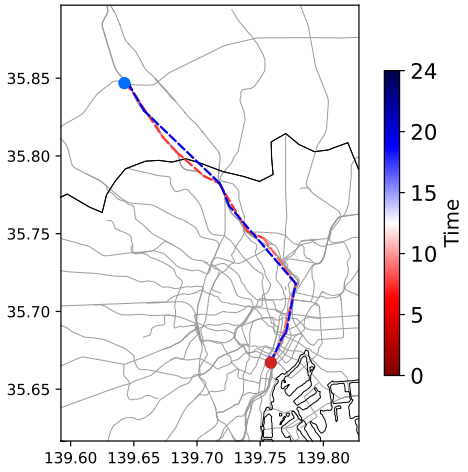


Figure S4: Comparison between monthly GPS-derived commuting residents and actual residential population. The grey dots show individual data points, while the red line with error bars represents binned averages with a slope of approximately 1.7. Two reference lines are shown: y = 1.7x (upper dotted line) and y = x (lower dotted line).



(a)



(b)

Figure S5: Map of a typical commuter's trajectory. Figures S5a and S5b illustrate a commuter's trajectory covering the entire Tokyo metropolitan area and its magnified view, respectively. Blue points on the map represent records marked as home status, whereas red points indicate works status records. The color of the trajectory lines signifies the time of day; this transition is indicated by the color bar, shifting from red (0 a.m.) to blue (23:59 p.m.). The trajectory map effectively captures the commuter's daily routine of leaving home in the morning and returning in the evening. This commuter's route overlaps with the shapes of railways, suggesting the potential use of underground transportation. The x- and y-axes correspond to latitude and longitude, respectively.

8 Energy cost

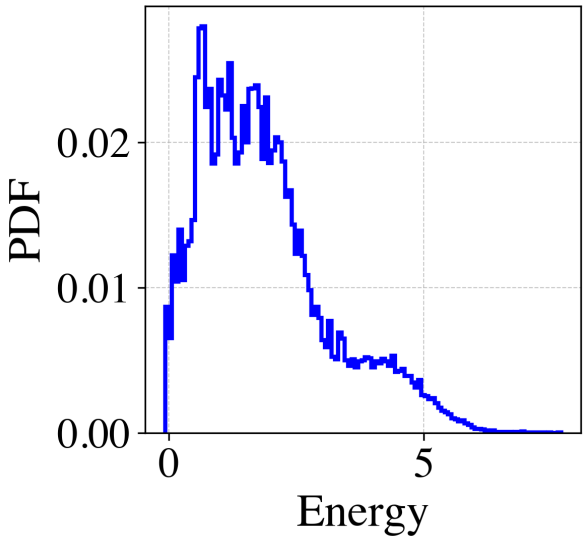


Figure S6: Probability Distribution Function (PDF) of Energy Cost. The distribution of energy costs for all identified route alternatives. The concentration towards lower energy values reflects the system's overall preference for efficient (low-cost) commuting options, with a diminishing tail for higher-cost routes.

9 Data accuracy

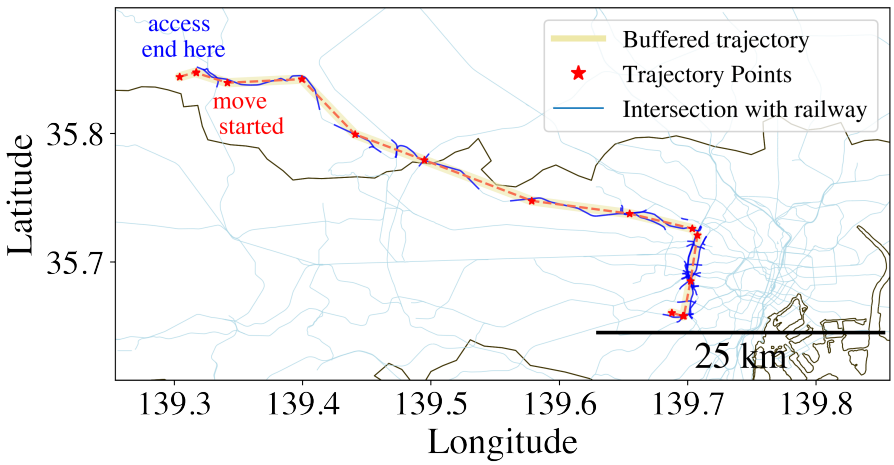


Figure S7

As illustrated in Fig. S7, the GPS logging frequency is insufficiently dense to pinpoint precise state changes; specifically, there is typically a spatial and temporal gap between the final point of the *access* walk and the first logged *move* point on the train. However, according to the data provider's (Agoop) handbook, the dataset reliably captures when a user initiates a trip (leaves home), changes movement states, or emerges from deep underground to transfer. This reliability stems from how the GPS satellite logs are triggered during state transitions

Based on the typical features of our dataset, we rely on features from the dataset robustly capture:

1. **Door-to-door commuting time:** This is calculated from the moment a user initiates a trip (leaves home) to the moment they reach their workplace.

2. **High-speed moving status:** Typical train-speed points allow us to match moving trajectories to the physical railway network.

3. **Possible transfer behaviors:** Low-speed *stroll* points captured between high-speed *move* segments successfully identify when a user is walking to transfer between lines.

10 Robustness Check

A minimum sample of 20 commuters might be insufficient to estimate 6 parameters, and could potentially explain the counterintuitive parameter distributions (heterogeneity). We address this from three perspectives to confirm that the observed heterogeneity is a genuine behavioral phenomenon rather than a small-sample artifact:

1. **Adequacy of the observation-to-parameter ratio:** In discrete choice modeling, the effective sample size entering the likelihood function is the number of choice situations (commuters × alternatives). Because we restrict our analysis to OD pairs with 2 to 6 route alternatives, even the most conservative scenario ($N = 20$ commuters, 2 alternatives) yields $20 \times 2 = 40$ observation rows for 6 parameters (see Table 17). This provides approximately 6.7 observations per parameter. Recent statistical studies [5] show that having 5 to 9 observations per parameter is enough to produce reliable and unbiased results, safely meeting the requirements for our model.

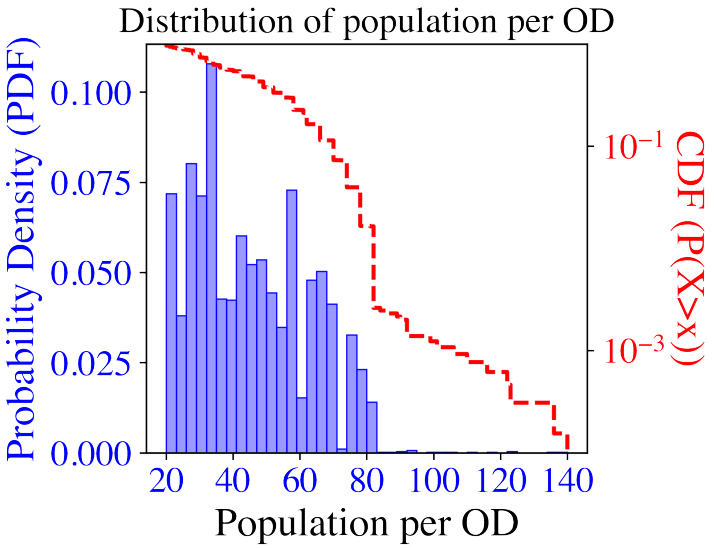
2. **Empirical distribution of OD-pair sample sizes:** The minimum threshold of 20 commuters is a conservative lower bound. As shown in Figure 8a, the distribution of per-OD sample sizes has a mode around 35 commuters, with a substantial right tail extending beyond 100. The complementary cumulative distribution function (CCDF, red dashed line) shows that OD pairs with 20 to 40 commuters account for only about half of the sampled ODs.

3. **Robustness check for behavioral heterogeneity:** To directly test if small samples drove the counterintuitive bimodal distributions of the Commuting Time (CT) and Commuting Distance (CD) coefficients, we performed a sensitivity analysis by re-estimating the models using a stricter threshold of $N \geq 40$ commuters. Even after restricting the data—which halves our available sample—the counterintuitive result that some ODs tend to prefer longer time and distance routes still exists, though the proportion turns lower. This confirms that small-sample OD pairs do not drive the reported heterogeneity patterns.

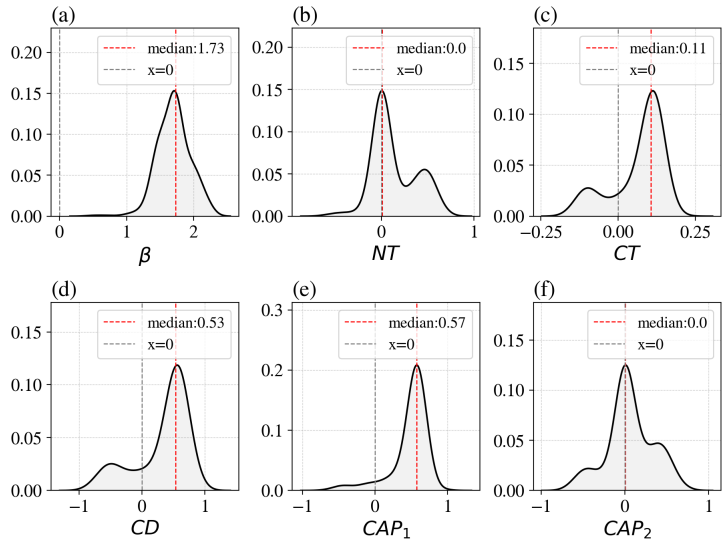
   Although ω_{NT} and ω_{CAP_2} show a tendency close to 0, this is because OD pairs with more than 40 commuters are geographically distributed closer to the urban center, naturally requiring fewer transfers. Additionally, these more central OD pairs contain fewer part-time and partial-telework commuters (who typically depart during off-peak hours), which mitigates the observed heterogeneity in the scale parameter (β).

Table 17: Illustration of the discret choice estimation data structure for a minimal OD pair (20 commuters, 2 route alternatives). Each commuter contributes two rows (one per alternative), yielding 40 observation rows for 6 parameters. Route attributes are median-aggregated values shared by all commuters within the same OD pair, meaning there is no within-route sampling noise in the explanatory variables.

Commuter n	Route j	CT (min)	log(CD) (km)	PK	NT	log(CAP_1)	Chosen y_{nj}
1	A	35	2.94	1	0	—	**1**
1	B	42	3.18	1	1	11.82	0
2	A	35	2.94	0	0	—	0
2	B	42	3.18	0	1	11.82	**1**
3	A	35	2.94	1	0	—	**1**
3	B	42	3.18	1	1	11.82	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
20	A	35	2.94	1	0	—	**1**
20	B	42	3.18	1	1	11.82	0
Total: 40 observation rows → 6 parameters to estimate							



(a) Distribution of per-OD pair sample sizes (number of commuters). Blue bars show the probability density function (PDF); red dashed line shows the complementary cumulative distribution function (CCDF, $P(X > x)$, log scale). The mode is around 25–30 commuters, with a smooth right tail extending beyond 140.



(b) The OD-Specific parameter estimation for ODs with more than 40 commuters.

Figure S8: Empirical distribution of sample sizes and the robustness check for parameter estimation.

11 Status Verification

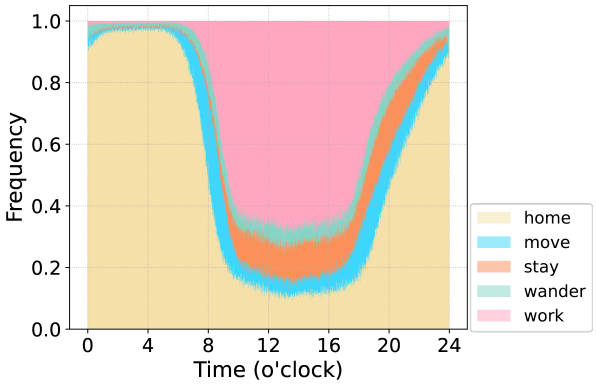


Figure S9: **Activity state patterns.** Temporal distribution of five activity states (home, move, stay, stroll, work) over 24 hours. Clear diurnal patterns emerge: home activity dominates nighttime (0-6, 20-24 hours), movement peaks during morning (7-9) and evening (17-19) rush hours, and work activity sustains during business hours (9-17).

12 CBD Trajectory Identification

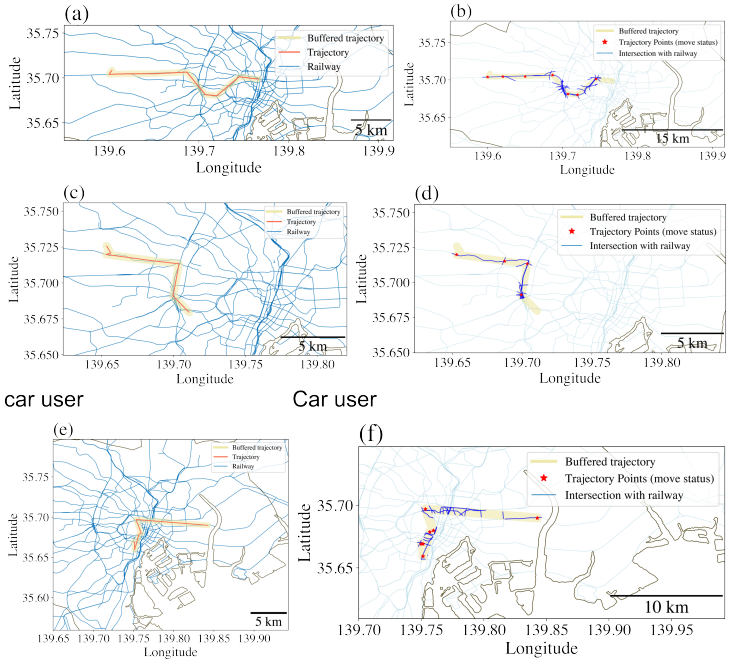


Figure S10: Example trajectories near the CBD area. (a)–(d): two railway commuters; (e)–(f): one car commuter whose moving points intersect with a substantially larger number of railway lines.

13 Express effect

We retain the original transfer count and introduce an explicit binary variable $ET_{nj} \in \{0, 1\}$ that identifies routes qualifying as express transitions (i.e., transfer routes that are faster than the fastest available direct route, per Eq. (9)):

$$ET_{nj} = \mathbf{1}\left[NT_{nj} > 0 \;\text{ and }\; CT_{nj} < \min_{k:\, NT_{nk}=0} CT_{nk}\right]. \tag{2}$$

The revised feature vector is:

$$X_{nj} = \big(CT_{nj},\ \log(CD_{nj}),\ PK_{nj},\ NT_{nj},\ ET_{nj},\ \log(CAP_{nj1}),\ \log(CAP_{nj2})\big). \tag{3}$$

To assess the effect of this change, we estimated three model specifications under the common-parameter (CPOD) setting (Table 18):

Table 18: Model comparison: treatment of express transitions

Specification	R^2	Hit Rate	AIC
Model A (original: adjusted NT, no ET)	0.232	55.8%	261,076
Model B (raw NT, no ET)	0.15	48.3%	393,630
Model C (raw NT + ET variable)	0.19	53.6%	446,759

After we add express as a route attribute, the estimation results of Model C are shown in Table 19.

Table 19: Estimation Results after add attribute *Express*

Parameter	**Estimate**	**Std. Error**	**t-value**
Scale parameter (β)	1.499	0.006	253.584
Number of Transfers	0.615^{***}	0.003	188.01
Commuting Time	0.091^{***}	0.022	89.662
Commuting distance	0.159^{***}	0.006	28.546
Peak Departure	-0.157^{***}	0.002	-98.729
Express	0.589^{**}	0.002	2.931
Capacity at 1st transfer	0.512^{*}	0.012	2.332
Capacity at 2nd transfer	-0.286^{***}	0.002	-166.767

AIC	446,759
R^2 (McFadden's R^2)	0.19
Hit rate	53.6%

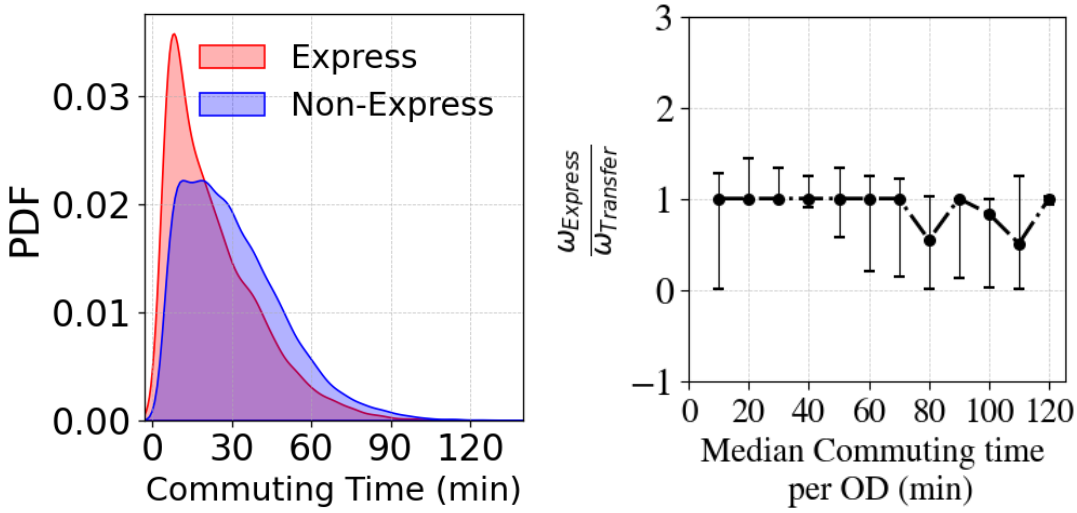
Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Model A (the original specification in the manuscript) achieves the highest ρ^2 and hit rate with the lowest AIC. However, Model B—which completely ignores the express effect—yields the lowest ρ^2. This confirms the reviewer's point: express transitions explain a significant portion of route choice behavior. Notably, in Model C the estimated weights for *Express* (0.589) and *Number of Transfers* (0.615) are similar in magnitude, suggesting that commuters assign nearly equal energy cost to express transitions and regular transfers.

We present two figures to elaborate on this finding. Fig. 11a shows the commuting time distributions for routes with and without express transitions. Express routes are associated with shorter commuting times, with a reduction of approximately 10 to 15 minutes.

To quantify the trade-off between express transitions and regular transfers, we computed the marginal rate of substitution $\omega^{(o)}_{\text{Express}}/\omega^{(o)}_{\text{NT}}$ for all OD pairs where both parameters are identified. Fig. 11b shows this trade-off ratio as a function of the median commuting time per OD pair.

For commutes shorter than approximately 70 minutes, the ratio is stably close to 1.0 with tight interquartile ranges. This indicates that commuters treat express transitions as carrying the same energy cost as regular transfers, the error bars shows some of ODs regard a express transfer less than 1 times transfer, and the length of error bar increases with the x-axis. For longer commutes (beyond 70 minutes), the ratio exhibits greater variability. This likely reflects heterogeneous preferences in this segment: some commuters on longer OD pairs may value the time savings from express service more highly, reducing the relative energy cost of the express transition, while others may place greater weight on uninterrupted seating and comfort.



(a) (b)

Figure S11

14 ODSP model's fitness

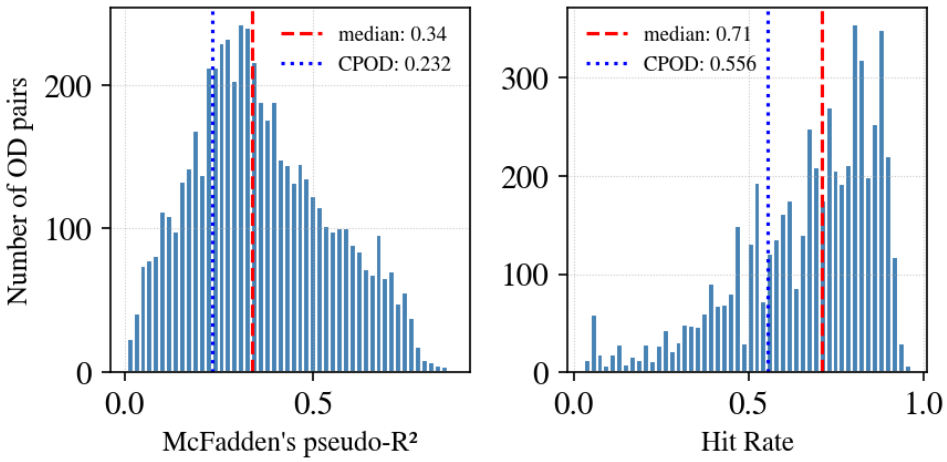


Figure S12: **Distribution of model fit statistics across OD-specific (ODSP) estimations.** (a) McFadden's pseudo-R^2 distribution across 5,908 OD pairs (median = 0.34). (b) Hit rate distribution (median = 0.71). Blue dotted lines indicate the corresponding CPOD aggregate model benchmarks from Table 6 ($R^2 = 0.232$, hit rate = 55.8%). The majority of OD-specific models outperform the aggregate benchmark: 74.8% of OD pairs exceed the CPOD R^2 and 73.8% exceed the CPOD hit rate, confirming that OD-level estimation captures local choice structure more effectively than the pooled model.

15 Non-railway commuters' trip interruption distribution

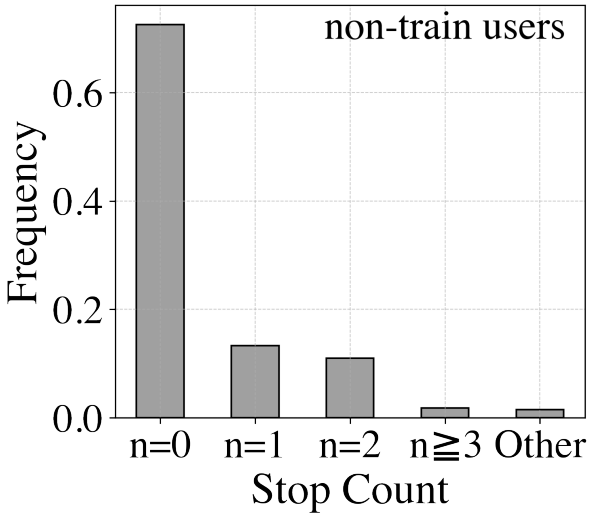


Figure S13: **Motif distributions for non-railway commuters.** Distribution of commuting motif types by number of interruptions (stop count) after transportation mode classification. Non-railway commuters (car, bus, walking, cycling): interruptions represent traffic stops, congestion, or brief pauses; direct routes (n=0) dominate at over 75%.

16 Maximum Entropy framework

The Maximum Entropy (MaxEnt) principle states that, given a set of known constraints, the least biased probability distribution is the one that maximizes Shannon entropy. We apply this principle to derive route choice probabilities under two different assumptions about spatial homogeneity.

Notation Let O denote the set of all Origin–Destination (OD) pairs, and let $o \in O$ index a specific OD pair. For each OD pair o, let N_o be the set of commuters and R_o be the number of available route alternatives, which varies across OD pairs. For commuter $n \in N_o$, let $j \in \{1, \ldots, R_o\}$ index the route alternatives. We define the following:

- P_{jn}: the probability that commuter n chooses route j,
- $\mathbf{X}_{jn} \in \mathbb{R}^M$: the attribute vector (e.g., travel time, number of transfers) of route j for commuter n,
- $y_{jn} \in \{0, 1\}$: the observed choice indicator (1 if commuter n chose route j, 0 otherwise).

Objective function: Shannon entropy For any system under consideration, we seek to maximize the total Shannon entropy across all commuters and route choices:

$$S = -\sum_{o \in O} \sum_{n \in N_o} \sum_{j=1}^{R_o} P_{jn} \ln P_{jn} \tag{4}$$

Model 1 (city-wide pooled estimation): Global model (single system, single parameter set) This method assumes *spatial homogeneity*: the entire metropolitan network constitutes a single macroscopic system. All commuters are assumed to share the same parameter vector $\boldsymbol{\omega}_{\text{raw}} \in \mathbb{R}^M$, regardless of which OD pair they belong to.
Normalization. For every commuter n in every OD pair o, the choice probabilities must sum to unity:

$$\sum_{j=1}^{R_o} P_{jn} = 1, \quad \forall\, n \in N_o,\ \forall\, o \in O \tag{5}$$

Global expected value constraint This equation enforces that the expected value of the route attributes (predicted by probabilities P_{jn}) must perfectly equal the empirically observed aggregate of those at-

tributes in the real world.

$$\sum_{o\in O}\sum_{n\in N_o}\sum_{j=1}^{R_o} P_{jn}\,\mathbf{X}_{jn} = \sum_{o\in O}\sum_{n\in N_o}\sum_{j=1}^{R_o} y_{jn}\,\mathbf{X}_{jn} \tag{6}$$

Derivation We construct the Lagrangian $\mathcal{L}$, introducing scalar multipliers α_{no} for the normalization constraints and a vector multiplier $\boldsymbol{\omega}_{\text{raw}} \in \mathbb{R}^M$ for the expected value constraint:

$$\mathcal{L} = -\sum_{o,n,j} P_{jn}\ln P_{jn} + \sum_{o,n}\alpha_{no}\left(1-\sum_{j=1}^{R_o} P_{jn}\right) + \boldsymbol{\omega}_{\text{raw}}^{\top}\left(\bar{\mathbf{X}}_{\text{obs}} - \sum_{o,n,j} P_{jn}\,\mathbf{X}_{jn}\right) \tag{7}$$

where $\bar{\mathbf{X}}_{\text{obs}}$ denotes the right-hand side of Eq. (6). Taking the partial derivative with respect to P_{jn} and setting it to zero:

$$\frac{\partial\mathcal{L}}{\partial P_{jn}} = -\big(\ln P_{jn}+1\big) - \alpha_{no} - \boldsymbol{\omega}_{\text{raw}}^{\top}\mathbf{X}_{jn} = 0 \tag{8}$$

Solving for P_{jn}:

$$P_{jn} = \exp\big(-1-\alpha_{no}-\boldsymbol{\omega}_{\text{raw}}^{\top}\mathbf{X}_{jn}\big) \tag{9}$$

Applying the normalization constraint (Eq. 5) to eliminate the constant $e^{-1-\alpha_{no}}$, we obtain:

$$\boxed{P_{jn} = \frac{\exp\big(-\boldsymbol{\omega}_{\text{raw}}^{\top}\mathbf{X}_{jn}\big)}{\displaystyle\sum_{j'=1}^{R_o}\exp\big(-\boldsymbol{\omega}_{\text{raw}}^{\top}\mathbf{X}_{j'n}\big)}} \tag{10}$$

Equivalence with maximum likelihood estimation. The vector $\boldsymbol{\omega}_{\text{raw}}$ arises here as a Lagrange multiplier enforcing the expected value constraint (Eq. 6). Equivalently, maximizing the log-likelihood $\ell = \sum_{o,n,j} y_{jn}\ln P_{jn}$ with P_{jn} given by Eq. (10) yields the first-order condition

$$\sum_{o,n,j} P_{jn}\,\mathbf{X}_{jn} = \sum_{o,n,j} y_{jn}\,\mathbf{X}_{jn} \tag{11}$$

which is identical to Eq. (6). Therefore the Lagrange multiplier and the MLE estimator satisfy the same equation and are numerically identical (Anas, 1983 [1]; Donoso & de Grange, 2010 [2]):

1. The Maximum Likelihood Estimator (MLE) Derivation In the standard Random Utility Maximization (RUM) framework, the probability P_{nj} that commuter n chooses route j is given by the Multinomial Logit (MNL) formula:

$$P_{nj} = \frac{\exp(-\boldsymbol{\omega}_{\text{raw}}^{\top}X_{nj})}{\sum_{k\in C_n}\exp(-\boldsymbol{\omega}_{\text{raw}}^{\top}X_{nk})}$$

Where X_{nj} is the vector of route attributes and $\boldsymbol{\omega}_{\text{raw}}$ is the unconstrained parameter vector to be estimated. The log-likelihood function to be maximized is:

$$\log\mathcal{L}_{\text{MLE}} = \sum_{n=1}^{N}\sum_{j\in C_n} y_{nj}\ln P_{nj}$$

To find the maximum likelihood, we calculate the score equation by taking the partial derivative of $\log\mathcal{L}_{\text{MLE}}$ with respect to $\boldsymbol{\omega}_{\text{raw}}$ and setting it to zero. First, we take the derivative of $\ln P_{nj}$:

$$\ln P_{nj} = -\boldsymbol{\omega}_{\text{raw}}^{\top}X_{nj} - \ln\sum_{k\in C_n}\exp(-\boldsymbol{\omega}_{\text{raw}}^{\top}X_{nk})$$

$$\frac{\partial\ln P_{nj}}{\partial\boldsymbol{\omega}_{\text{raw}}} = -X_{nj} - \frac{\sum_{k\in C_n}\exp(-\boldsymbol{\omega}_{\text{raw}}^{\top}X_{nk})(-X_{nk})}{\sum_{k\in C_n}\exp(-\boldsymbol{\omega}_{\text{raw}}^{\top}X_{nk})}$$

Recognizing the MNL probability formula in the second term, this simplifies to:

$$\frac{\partial \ln P_{nj}}{\partial \boldsymbol{\omega}_{\text{raw}}} = -X_{nj} + \sum_{k \in C_n} P_{nk} X_{nk}$$

Now, substitute this back into the derivative of the full log-likelihood function:

$$\frac{\partial \log \mathcal{L}_{\text{MLE}}}{\partial \boldsymbol{\omega}_{\text{raw}}} = \sum_{n=1}^{N} \sum_{j \in C_n} y_{nj} \left(-X_{nj} + \sum_{k \in C_n} P_{nk} X_{nk} \right) = 0$$

Distribute the y_{nj} term:

$$-\sum_{n=1}^{N} \sum_{j \in C_n} y_{nj} X_{nj} + \sum_{n=1}^{N} \sum_{j \in C_n} y_{nj} \sum_{k \in C_n} P_{nk} X_{nk} = 0$$

Because each commuter n chooses exactly one route, the sum of their choices $\sum_{j \in C_n} y_{nj} = 1$. The equation collapses to:

$$-\sum_{n=1}^{N} \sum_{j \in C_n} y_{nj} X_{nj} + \sum_{n=1}^{N} \left((1) \sum_{k \in C_n} P_{nk} X_{nk} \right) = 0$$

Changing the dummy index k back to j allows us to combine the terms, yielding the final MLE first-order condition:

$$\sum_{n=1}^{N} \sum_{j \in C_n} X_{nj} \left[y_{nj} - P_{nj}(\boldsymbol{\omega}_{\text{raw}}) \right] = 0$$

2. The Maximum Entropy (MaxEnt) Derivation The MaxEnt approach makes no assumptions about utility maximization. Instead, it seeks the least biased probability distribution P_{nj} by maximizing Shannon entropy S, subject to known macroscopic constraints.

$$S = -\sum_{n=1}^{N} \sum_{j \in C_n} P_{nj} \ln P_{nj}$$

Constraint 1 (Normalization): The probabilities for each commuter must sum to 1.

$$\sum_{j \in C_n} P_{nj} = 1 \quad \forall n$$

Constraint 2 (Expected Value): The model's expected macroscopic attribute costs must perfectly equal the empirically observed attribute costs.

$$\sum_{n=1}^{N} \sum_{j \in C_n} P_{nj} X_{nj} = \sum_{n=1}^{N} \sum_{j \in C_n} y_{nj} X_{nj}$$

We construct the Lagrangian $\mathcal{L}_{\text{MaxEnt}}$ using scalar multipliers α_n for the normalization constraints and a vector multiplier $\boldsymbol{\omega}_{\text{raw}}$ for the expected value constraint:

$$\mathcal{L}_{\text{MaxEnt}} = -\sum_{n,j} P_{nj} \ln P_{nj} + \sum_{n} \alpha_n \left(1 - \sum_{j} P_{nj} \right) + \boldsymbol{\omega}_{\text{raw}}^{\top} \left(\sum_{n,j} y_{nj} X_{nj} - \sum_{n,j} P_{nj} X_{nj} \right)$$

To find the maximum entropy distribution, we take the partial derivative with respect to P_{nj} and set it to zero:

$$\frac{\partial \mathcal{L}_{\text{MaxEnt}}}{\partial P_{nj}} = -\ln P_{nj} - 1 - \alpha_n - \boldsymbol{\omega}_{\text{raw}}^{\top} X_{nj} = 0$$

Solving for P_{nj}:

$$P_{nj} = \exp(-1 - \alpha_n) \exp(-\boldsymbol{\omega}_{\text{raw}}^{\top} X_{nj})$$

Applying Constraint 1 (Normalization) to eliminate the constant $\exp(-1-\alpha_n)$, we recover the exact MNL functional form:

$$P_{nj} = \frac{\exp(-\boldsymbol{\omega}_{\text{raw}}^\top X_{nj})}{\sum_{k\in C_n}\exp(-\boldsymbol{\omega}_{\text{raw}}^\top X_{nk})}$$

Finally, taking the partial derivative of the Lagrangian with respect to the vector multiplier $\boldsymbol{\omega}_{\text{raw}}$ yields the stationarity condition, which simply returns the original constraint:

$$\frac{\partial \mathcal{L}_{\text{MaxEnt}}}{\partial \boldsymbol{\omega}_{\text{raw}}} = \sum_{n=1}^{N}\sum_{j\in C_n} y_{nj}X_{nj} - \sum_{n=1}^{N}\sum_{j\in C_n} P_{nj}X_{nj} = 0$$

Rearranging this gives:

$$\sum_{n=1}^{N}\sum_{j\in C_n} X_{nj}\big[y_{nj} - P_{nj}(\boldsymbol{\omega}_{\text{raw}})\big] = 0$$

The MLE score equation and the MaxEnt stationarity condition resolve to the exact same algebraic formula. The parameter vector $\boldsymbol{\omega}_{\text{raw}}$, whether interpreted econometrically as the coefficients maximizing the likelihood of individual choices, or interpreted physically as the Lagrange multipliers enforcing aggregate systemic behavior, mathematically serves the identical function.

In practice, we estimate $\boldsymbol{\omega}_{\text{raw}}$ via standard maximum likelihood; the Maximum Entropy derivation above provides the theoretical justification for the logit functional form.

The generalized cost is $E_{jn} = \boldsymbol{\omega}_{\text{raw}}^\top \mathbf{X}_{jn}$. The parameter vector $\boldsymbol{\omega}_{\text{raw}}$ can be decomposed into a scalar sensitivity (inverse temperature) $\beta = \|\boldsymbol{\omega}_{\text{raw}}\|_2$ and a unit-weight direction vector $\boldsymbol{\omega} = \boldsymbol{\omega}_{\text{raw}}/\beta$, so that $\boldsymbol{\omega}_{\text{raw}} = \beta\,\boldsymbol{\omega}$ with $\|\boldsymbol{\omega}\|_2 = 1$. Under this decomposition:

$$P_{jn} = \frac{\exp\big(-\beta\,\boldsymbol{\omega}^\top \mathbf{X}_{jn}\big)}{\displaystyle\sum_{j'=1}^{R_o}\exp\big(-\beta\,\boldsymbol{\omega}^\top \mathbf{X}_{j'n}\big)} \tag{12}$$

Here $\boldsymbol{\omega}$ captures the relative preferences among route attributes and β controls the overall cost sensitivity. The parameters are *global*: they represent city-wide average preferences and cost sensitivity.

Model 2: OD-specific model ($|O|$ sub-systems, $|O|$ parameter sets) This method assumes *spatial heterogeneity*: the metropolitan network is decomposed into $|O|$ isolated sub-systems, one per OD pair. The population traveling within each OD pair o may exhibit distinct cost sensitivities and attribute trade-offs.

The normalization constraint is identical to the global model (Eq. 5). The critical difference lies in the expected value constraint: for each OD pair o, the model-predicted expected attribute vector must match the observed average *within that OD pair*. This yields $|O|$ vector constraints:

$$\sum_{n\in N_o}\sum_{j=1}^{R_o} P_{jn}^o\,\mathbf{X}_{jn} = \sum_{n\in N_o}\sum_{j=1}^{R_o} y_{jn}\,\mathbf{X}_{jn}, \quad \forall\, o\in O \tag{13}$$

Derivation Because the sub-systems are independent, the total entropy (Eq. 4) decomposes additively: $S = \sum_{o\in O} S^o$, where

$$S^o = -\sum_{n\in N_o}\sum_{j=1}^{R_o} P_{jn}^o \ln P_{jn}^o \tag{14}$$

This additive decomposition permits independent optimization of each sub-system. The Lagrangian for OD pair o is:

$$\mathcal{L}^o = -\sum_{n\in N_o}\sum_{j=1}^{R_o} P_{jn}^o \ln P_{jn}^o + \sum_{n\in N_o}\alpha_n\left(1-\sum_{j=1}^{R_o}P_{jn}^o\right) + \big(\boldsymbol{\omega}_{\text{raw}}\big)_o^\top\left(\bar{\mathbf{X}}_{\text{obs}}^o - \sum_{n\in N_o}\sum_{j=1}^{R_o} P_{jn}^o\,\mathbf{X}_{jn}\right) \tag{15}$$

where $(\boldsymbol{\omega}_{\text{raw}})_o \in \mathbb{R}^M$ is the OD-specific vector Lagrange multiplier. Taking the partial derivative with respect to P_{jn}^o and setting it to zero:

$$P_{jn}^o = \exp\bigl(-1 - \alpha_n - (\boldsymbol{\omega}_{\text{raw}})_o^\top \mathbf{X}_{jn}\bigr) \tag{16}$$

Applying the normalization constraint to eliminate α_n:

$$\boxed{P_{jn}^o = \frac{\exp\bigl(-(\boldsymbol{\omega}_{\text{raw}})_o^\top \mathbf{X}_{jn}\bigr)}{\sum_{j'=1}^{R_o} \exp\bigl(-(\boldsymbol{\omega}_{\text{raw}})_o^\top \mathbf{X}_{j'n}\bigr)}} \tag{17}$$

The same entropy–logit duality applies: $(\boldsymbol{\omega}_{\text{raw}})_o$ is estimated via maximum likelihood for each OD pair independently.

Analogously, we decompose $(\boldsymbol{\omega}_{\text{raw}})_o = \beta^o \boldsymbol{\omega}^o$, where $\beta^o = \|(\boldsymbol{\omega}_{\text{raw}})_o\|_2$ and $\|\boldsymbol{\omega}^o\|_2 = 1$:

$$P_{jn}^o = \frac{\exp\bigl(-\beta^o (\boldsymbol{\omega}^o)^\top \mathbf{X}_{jn}\bigr)}{\sum_{j'=1}^{R_o} \exp\bigl(-\beta^o (\boldsymbol{\omega}^o)^\top \mathbf{X}_{j'n}\bigr)} \tag{18}$$

The parameters $(\boldsymbol{\omega}_{\text{raw}})_o$ (equivalently β^o and $\boldsymbol{\omega}^o$) are *local*: they characterize the thermodynamic state—the specific preferences and cost sensitivities—of the sub-population within OD pair o.

References

[1] Anas, A. Discrete choice theory, information theory and the multinomial logit and gravity models. *Transportation Research Part B: Methodological 17*, 1 (1983), 13–23.

[2] Donoso, P., and de Grange, L. A microeconomic interpretation of the maximum entropy estimator of multinomial logit models and its equivalence to the maximum likelihood estimator. *Entropy 12*, 10 (2010), 2077–2084.

[3] Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. *Multivariate Data Analysis*, 7th ed. Pearson Prentice Hall, Upper Saddle River, NJ, 2009.

[4] Ministry of Land, Infrastructure, Transport and Tourism (MLIT). Transfer survey report 2017, 2017. [In Japanese].

[5] Vittinghoff, E., and McCulloch, C. E. Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology 165*, 6 (2007), 710–718.