



Article

Do Railway Commuters Exhibit Consistent Route Choice “Rationality” Across Different Contexts and Time? Evidence from Tokyo Metropolitan Commutes

Yixuan Y. Zheng¹, Hideki Takayasu² and Misako Takayasu^{1,2,*}¹ Department of Systems and Control Engineering, Institute of Science Tokyo, Yokohama 2268502, Japan² Department of Computer Science, Institute of Science Tokyo, Yokohama 2268502, Japan* Correspondence: takayasu@comp.isct.ac.jp**How To Cite:** Zheng, Y.Y.; Takayasu, H.; Takayasu, M. Do Railway Commuters Exhibit Consistent Route Choice “Rationality” Across Different Contexts and Time? Evidence from Tokyo Metropolitan Commutes. *Journal of Social Physics* 2026, 1(1), 6.

Received: 14 October 2025

Revised: 18 May 2026

Accepted: 10 June 2026

Published: 24 June 2026

Abstract: In urban railway systems, if every commuter were perfectly rational, concentrated demand on optimal routes would undermine system performance. Understanding the actual degree of collective route choice determinism, and whether it remains stable over time, is essential for transportation policy. Smartphone GPS data offers both the scale and complete trajectory coverage that survey data and smart card data individually lack, yet its coarse spatial accuracy has limited its adoption for railway route identification. To overcome this, we develop a methodology to identify railway commuting routes from one year of GPS data covering over one million daily users in the Tokyo metropolitan area. We apply a multinomial logit (MNL) framework, utilizing a standardization convention to extract a comparative measure of collective determinism alongside relative attribute preferences. We find that collective route choices in Tokyo are strongly cost-sensitive but non-deterministic. Both components remain stable across all twelve months of 2023, confirming that this determinism represents a structural consistency of the commuting system. Leveraging the dataset’s scale, we further estimate parameters for each origin–destination (OD) pair individually, revealing systematic heterogeneity driven by departure time and transport complexity that smaller datasets treat as unobserved.

Keywords: discrete choice modeling; collective behavior; human mobility; smartphone GPS data

1. Introduction

In urban railway systems, a fundamental tension exists between individual cost-minimization and collective system performance. If every commuter were perfectly rational, always selecting the route that minimizes their personal cost, demand would concentrate on a small number of optimal routes, degrading service quality and undermining system resilience. Conversely, if choices were entirely random, infrastructure investments designed around predictable demand patterns would lose their effectiveness. The actual degree of choice determinism, where the collective population falls on this spectrum between perfect predictability and pure randomness, therefore carries direct consequences for capacity planning, congestion management, and disruption response [1–4]. Equally important is whether this behaviour determinism level remains stable over time, or fluctuates with seasons, service changes, and external shocks. A system operating in a consistent state is fundamentally different, from a policy perspective, than one exhibiting volatile behavioral shifts.

Quantifying collective determinism, however, imposes a demanding requirement on data. Because the quantity of interest is a population-level parameter describing the distribution of choices across both cost-driven and idiosyncratic commuters, the dataset must be large enough to capture the full spectrum of decision-making behavior. Small samples risk overrepresenting particular behavioral segments and producing unstable parameter estimates. Moreover, to detect whether choice determinism varies systematically across contexts, such as departure times,



commuting distances, or transfer complexity, the dataset must support disaggregate analysis at the level of individual origin-destination (OD) pairs while retaining sufficient observations within each pair for statistical reliability.

Traditional route choice research [5–13], grounded in the Random Utility Maximization (RUM) framework and its extensions such as the Mixed Logit and Latent Class models [14–16], has relied primarily on stated or revealed preference survey data involving tens to hundreds of respondents [5, 9, 15, 17–19]. While these models have produced foundational insights into route attribute trade-offs, their sample sizes fundamentally limit the ability to observe population-level phenomena or to estimate parameters for individual OD pairs. Heterogeneity in preferences must be pre-specified as a parametric distribution [14–16] (continuous in Mixed Logit [20], discrete in Latent Class [21, 22]) rather than empirically observed from the data [5, 9, 15, 18, 20, 23]. This constraint means that the core question - what is the actual level of collective determinism in a metropolitan commuting system, has remained largely unaddressed.

Larger-scale passively collected data sources have begun to relax this constraint. Public transport smart card data (e.g., Suica, Oyster) provides transaction records for millions of trips, but captures only entry and exit stations [8, 24–26]. Crucially, transfers between lines operated by the same company do not require re-tapping the card, rendering these transfers invisible to the analyst [24, 25]. Since transfer behavior [6, 13] is a primary source of route differentiation in dense networks like Tokyo’s, this limitation directly undermines route identification accuracy. Smartphone GPS trajectory data offers a complementary advantage: it records complete door-to-door journeys at high temporal resolution, capturing the actual paths taken including all intermediate transfers. The dataset used in this study covers approximately 1.2 million anonymous users per day across the Tokyo metropolitan area (Shutoken) throughout 2023, providing the scale and trajectory completeness that the research question demands. However, GPS data presents its own methodological challenge: the coarse spatial accuracy of smartphone positioning (approximately 10 m on average) has limited its adoption for railway route identification [7, 24, 27–29], where distinguishing between routes that share station proximity requires careful spatial analysis [7].

To address this challenge, we develop a comprehensive methodology for extracting commuting trajectories, transportation modes, and transfer stations from large-scale GPS data. Our approach combines rule-based transport mode classification validated against national statistics with a Voronoi tessellation-based spatial matching technique that identifies transfer stations from GPS trajectory interruptions. This pipeline transforms raw location records into structured route choice observations suitable for discrete choice modeling without requiring full path enumeration over the railway network graph.

For the modeling framework, we adopt the Multinomial Logit (MNL) model [14–16]. In standard Random Utility Maximization (RUM) applications, the scale parameter is typically normalized to one, with its role in reflecting the error term’s variance absorbed into the coefficient magnitudes [15]. In this study, we additionally report the norm and direction of the estimated coefficient vector as descriptive summaries; we apply a standardization convention that constrains the ℓ_2 -norm of the utility weights to unity. Under this shared scale, we extract a comparative magnitude β alongside the unit-length direction ω . This convention serves a specific analytical purpose: β captures the overall sensitivity of collective route choices to cost differences, serving as a macroscopic measurement of collective determinism and quantifying how deterministic or random the population’s behavior is; while ω captures the relative importance of each route attribute. Furthermore, because our dataset covers over one million daily users, the sample size is sufficient to estimate β and ω not only at the city-wide level but also for individual OD pairs, allowing systematic heterogeneity to be directly observed from the data rather than pre-specified as a parametric distribution as in the Mixed Logit framework [20].

In this paper, we focus on quantifying choice determinism, the degree to which collective behavior is predictable based on observed costs. Although we refer to this as “rationality” [30] in the title, we employ the term strictly as a macroscopic metaphor. It describes the ensemble’s tendency to concentrate on optimal routes rather than choosing randomly, without implying any behavioral assumptions about individual commuter intent.

Applying this framework to our dataset, we obtain several findings. First, Tokyo railway commuters’ choices exhibit strong but non-deterministic concentration: collective choices strongly but not absolutely favor lower-cost routes, and this level remains stable across all twelve months of 2023, confirming that the commuting system operates in a structural consistency. Second, leveraging the dataset’s scale, we estimate parameters for each OD pair individually, revealing systematic heterogeneity driven by departure time and transport complexity. Peak-hour commuters exhibit less deterministic and more strategic, exploratory behavior than off-peak commuters, reflecting behavioral adaptation under capacity constraints. Third, we identify transfers as the dominant source of choice uncertainty. We quantify the transfer-equivalent time penalty as ranging from approximately 7 min for short-to-medium commutes to over 20 min for long-distance suburban routes, demonstrating that these transfer-induced interruptions create bimodal travel time distributions that fundamentally limit model predictability for transfer-intensive routes. These OD-specific analyses are enabled precisely by the unprecedented data scale; with smaller

datasets, the per-OD sample sizes would be insufficient for reliable estimation, and this heterogeneity would be absorbed into unobserved random terms.

The remainder of this paper is organized to reflect a continuous analytical pipeline, moving from raw trajectory data to multi-scale behavioral insights, as illustrated in Figure 1. Section 2 establishes the empirical foundation by introducing the large-scale smartphone GPS dataset. To overcome the coarse spatial resolution of this data, Section 3 details our methodology for translating raw coordinates into observable behavior, utilizing transport mode classification and commuting interruption analysis to construct reliable choice sets. Building on this, Section 4 formulates the theoretical route choice model within a standard Multinomial Logit (MNL) framework, expressing decision-making in terms of collective determinism (quantified by the scale parameter, β) and relative attribute preferences. Section 5 presents our empirical findings through a top-down approach: first evaluating macro-level, system-wide consistency, and subsequently uncovering micro-level systematic heterogeneity across individual origin-destination pairs. Finally, Section 6 synthesizes these structural and behavioral findings, concluding with implications for urban transportation policy.

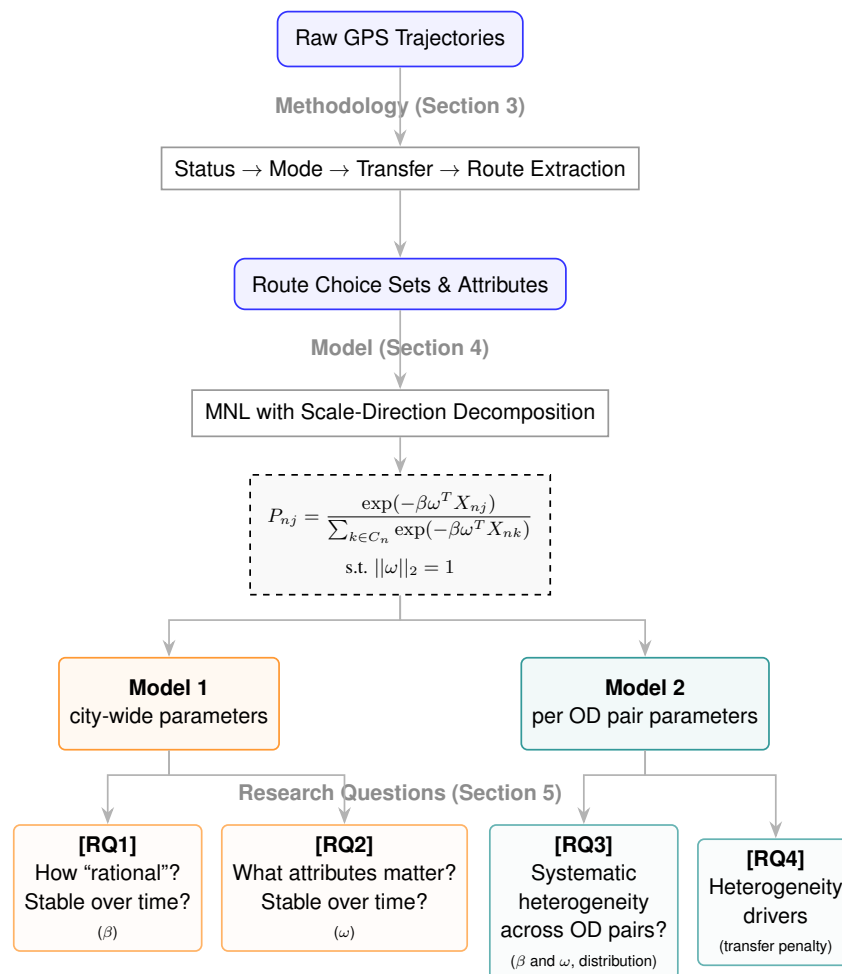


Figure 1. Conceptual framework of this study. The analytical pipeline proceeds from raw GPS trajectories through methodology (Section 3) to the MNL route choice model with scale-direction decomposition (Section 4). The model is estimated in two variants: Model 1 (city-wide pooled estimation) and Model 2 (OD-specific estimation), which together address four research questions on collective determinism, temporal stability, and systematic heterogeneity (Section 5).

Comparing with Previous Research

Table 1 provides a structured comparison with representative route choice studies. Three dimensions distinguish the current study. First, in data scale: previous GPS-based studies [24,29] tracked 35–156 users, while survey-based studies [17,18] covered at most several thousand respondents. Our dataset comprises approximately 1.2 million daily users, providing sufficient observations for OD-pair-level estimation. Second, in route identification: rather than generating synthetic alternatives via network algorithms (e.g., BFS-LE) or hypothetical survey designs, we construct choice sets directly from observed GPS traces by identifying transfer station sequences. Third, in variable derivation: transfer penalties and peak-hour effects are estimated from revealed behavior rather than stated preference scenarios.

Table 1. Comparison of route choice models in previous studies.

Item	Tang et al. [24]	Ma et al. [17]	Montini et al. [29]	Garcia-Martinez et al. [18]	This Study
Data Type	GPS and GIS data; survey data	2011 Household Travel Survey	GPS tracking, OpenStreetMap, elevation, and survey data	RP and SP survey data	GPS trajectory data with land use and station-level attributes
Sample Size	35 individuals (109 trips)	2967 samples from 29,670 trips by 9100 participants	156 valid users	78 respondents; 295 valid answers	1.2 million users, per user per path choice
Route Choice Condition	Fixed ODs; observed GPS over multiple commuting days	Mode-time combinations (not spatial routes), based on survey	Observed GPS; ODs vary; includes multiple trip purposes	Routes defined by survey design	Fixed ODs; commuter inferred from status algorithm
Model Used	Neural Network (ANN)	Nested Logit (time-mode hierarchy)	MNL and Path Size Logit (joint mode-route)	Error Component Logit (ECL)	MNL with GPS-derived transfer penalties
Alternative Routes	10 route options per commuter generated using an extended BFS-LE algorithm (includes observed, shortest, and 8 synthetic routes based on network and traffic data)	Predefined combinations of mode and travel time bands; not actual physical routes	BFS-LE used for car/bike/walk; via-point and stop-based methods for public transport	Hypothetical public transport trips with 0, 1, or 2 transfers; generated via Ngene software in SP survey	Observed choice sets from GPS traces; alternatives identified by transfer patterns
Variables Used	Traveler classification used demographics, job type, and household information. Route choice model included route length, detour rate, freeway access, number of intersections, and left turns	Mode-time choice modeled using demographics (age, income, job), household vehicle/bike ownership, and commute distance	Route utility based on travel time, road type ratio, elevation, number of turns/signals, trip purpose, and environmental attitudes	Utility modeled using transfer count, walking/waiting time, crowding, stairs, and user characteristics (e.g., age, gender, habits)	Transfer count, commuting time, peak hour dummy, station capacity; variables derived from real movement data
Results	ANN classification accuracy: 79%; route choice prediction accuracy: 67.3%	Model fit: $R^2 = 0.428/0.111$; hit rate: 86.3%	PT model: $R^2 = 0.541$ (Basic CSG), 0.595 (Via CSG); joint model: $R^2 = 0.326$	No fit index reported; perceived transfer penalty estimated at approximately 15–18 in-vehicle minutes	$R^2 = 0.232$; hit rate $\approx 55.8\%$; Spearman $\rho = 0.72$ (city-wide), 0.86 (OD-specific); parameters stable across 12 months

2. Data Description and Preprocessing

This and next sections outline our approach to acquiring, processing, and refining large-scale GPS trajectory data for commuting pattern analysis. In this section we describe the primary dataset (Section 2.1), which provides high-resolution spatiotemporal mobility information from smartphone across Japan's Shutoken metropolitan area. We then detail our first-stage data processing (Section 2.2), including filtering criteria and population renormalization techniques to address sampling biases.

2.1. Data Description

This study utilizes a large-scale mobility GPS dataset provided by a private company, Agoop Corporation, containing anonymized location data from approximately 1.2 million smartphones per day across Japan, with each user having an average of 125 trajectory points per weekday [31, 32]. The GPS data provide timestamped latitude-longitude coordinates with high temporal resolution, typically captured at one-minute intervals, and an average spatial accuracy of approximately 10 m. Each data point includes a randomized user ID (reset nightly to protect privacy), timestamp, geographical coordinates, and associated home and work city codes. Additionally, we utilize railway network Geographic Information System (GIS) data from the Ministry of Land, Infrastructure, Transport and Tourism [33], which contains station locations (coordinates and attributes), railway line geometries (polyline representations), and transfer station indicators for transportation mode identification and route matching.

To ensure robust analysis of regular commuting patterns, we focused exclusively on weekday data throughout 2023 for the Shutoken area of Japan (Tokyo, Kanagawa, Chiba, Saitama), excluding weekends and holidays. This study period and area allowed us to concentrate on typical workday route choice patterns that represent routine commuting behavior.

We should note that the dataset has certain unavoidable limitations. First, as it relies on smartphone applications issued by a private company, there is a demographic bias in representation. According to the Ministry of Internal Affairs and Communications data, elderly individuals and children under 13 are underrepresented due to lower smartphone adoption rates in these groups [31]. Second, privacy protections result in certain data limitations—user IDs are randomized each night, preventing multi-day trajectory analysis for a single individual, and precise residential locations are blurred to central points within grid areas.

2.2. Data Preprocessing

To prepare this extensive dataset for commuting pattern analysis, we implemented several data preprocessing steps to ensure quality and relevance. Several filtering criteria were applied to focus specifically on commuting behavior:

- Only users with more than 100 location points per day were included to ensure sufficient trajectory information.
- Selected IDs where both home city code and work city code were within the study area.
- Applied population renormalization to align the GPS user sample with actual population figures based on official statistics, where the real population is consistently around 1.7 times the recorded monthly number of our dataset, as shown in Supplementary Figure S4.

3. Methodology

The model formulated in Section 4 requires four quantities per recorded user in the GPS data: whether the user is a railway commuter, the location information of their origin and destination (OD pair), the sequence of transfer stations defining their chosen route, and the attributes characterizing each route alternative. This section describes how each quantity is extracted from raw GPS trajectories.

We first classify every GPS point into an activity status—*home*, *work*, *move*, *stroll*, *stay*—which provides the building blocks for everything that follows (Section 3.1). These status labels allow us to locate each user's home and work to define the OD pair, extract the morning commute segment between them, and distinguish short pauses (*stroll*) from prolonged stops (*stay*).

Next, we classify each commuter's primary transport mode—railway, car or bus, bicycle, walking—and restrict the subsequent analysis to railway users (Section 3.2).

Finally, from each railway commuter's trajectory we extract *interruptions*, defined as *stroll* states bounded by *move* states on both sides. An interruption occurring within a station-area grid cell is identified as a transfer (Section 3.3).

3.1. Status Identification

Based on previous research [31, 32, 34], we further developed an algorithm to classify each point within the trajectory data to detailed movement states (*home*, *work*, *move*, *stroll*, *stay*), according to Table 2. Let $S(t)$ represent the activity status of a commuter at time t , defined as

$$S(t) \in \{home, work, move, stroll, stay\} \quad (1)$$

The complete trajectory can therefore be expressed as an ordered sequence:

$$\{S(t_1), S(t_2), \dots, S(t_n)\} \quad \text{where } t_1 < t_2 < \dots < t_n \quad (2)$$

This classification forms the foundation for our subsequent analysis of commuting patterns. The time-based status distribution is shown in Supplementary Figure S9.

Table 2. Criteria for movement state classification.

Status	Classification Criteria
Home	Located within the identified 100m residential grid cell; stay duration exceeding 4 h; first observation after 5:00 AM
Work	Located within the identified 100m workplace grid cell; stay duration exceeding 5 h; not coinciding with home location
Move	Speed exceeding 8 km/h ¹ ; or stopping time less than 4 min ² within two consecutive mesh cells
Stroll	Speed below 8 km/h; stopping time exceeding 4 min ² ; continuous duration in a 1 km grid cell for less than 30 min
Stay	Speed below 8 km/h; continuous duration in a 1 km grid cell exceeding 30 min

¹ The 8 km/h threshold (approximately 2.2 m/s) exceeds typical walking speed (1.3 m/s [35]) and approximates adult running speed, ensuring that the *move* state captures active transport (by any mode) rather than pedestrian movement. ² Pauses shorter than 4 min are treated as part of the *move* status because they typically reflect incidental halts within an ongoing trip rather than a distinct visit. Pauses exceeding this threshold are more likely to correspond to structurally meaningful interruptions; the 4-min cutoff is calibrated against the average interruption duration reported in [36].

Then we represent a user's complete daily trajectory \mathcal{T}' as a sequence of tuples $(p_i, t_i, S(t_i))$ of i -th GPS location p_i , timestamp t_i , and user's status $S(t_i)$. We then extract the morning commute segment as all trajectory points between home departure and work arrival, by identifying the departure time $t_{\text{departure}}$ (time of the first non-home status after being at home) and arrival time t_{arrival} (time of the first work status after leaving home), and then calculate the door-to-door commuting time T_{commute} and commuting distance D_{commute} :

$$T_{\text{commute}} := t_{\text{arrival}} - t_{\text{departure}} \quad (3a)$$

$$D_{\text{commute}} := \sum_{i=1}^{M-1} d(p_i, p_{i+1}), \quad \text{for all } p_i \text{ where } t_{\text{departure}} \leq t_i \leq t_{\text{arrival}} \quad (3b)$$

where M is the number of GPS points recorded during the commute period, and $d(p_i, p_{i+1})$ is the geographic distance between consecutive points calculated using the Haversine formula to account for Earth's curvature.

For visualization purposes, Supplementary Figure S5 illustrates the daily trajectory of a typical commuter, showcasing the complete movement pattern: departure from the *home* location in the morning, extended stay at the *work* location during daytime hours, and return journey to the *home* location in the evening.

Each of the five states plays a distinct role in the downstream pipeline. *home* and *work* locate the OD pair and bound the morning commute segment (this section); *move* enables transport mode classification through GIS proximity (Section 3.2); also *move*, *stroll* and *stay* are the raw input to interruption and transfer detection (Section 3.3) and excludes non-commuting trajectories.

3.2. Transportation Mode Identification

The commuting motifs provide valuable insight into the structural patterns of daily journeys; however, understanding transportation mode choice is essential for interpreting these motifs within the broader context of urban mobility systems [29,37–40]. During a commuting process, users may employ multiple combinations of transportation modes when selecting their routes [38,41]. To address this complexity, we developed a comprehensive approach to identify each user’s primary mode of transport, defined as the mode accounting for the largest proportion of total travel distance among all trip segments.

We then constructed an algorithm to classify commuting trajectories into four transportation modes, railway, walking, cycling, bus or car, by extracting sequence-based features of velocity, acceleration, and spatial patterns from GPS trajectories, partially following the methodologies of [28,42]. Since cars and buses share the same road network and distinguishing them requires rather high-resolution data [38,43,44], they are combined into a single category in this study. The classification procedure follows a rule-based framework, as detailed in Table 3.

Table 3. Criteria for transportation mode classification.

Mode	Classification Criteria
Walking	Mean speed below 6 km/h; Maximum speed below 12 km/h; Total travel distance less than 4 km
Cycling	Mean speed below 18 km/h; Maximum speed below 30 km/h; Total travel distance less than 10 km
Railway	Travel distance greater than 1 km; Average moving points’ distance to railway network less than 90 m ¹
Bus or Car	All remaining trajectories after other mode classification

¹ Railway classification incorporates GIS-based proximity analysis.

Among these classification methods, railway identification required a particularly sophisticated approach due to the unique characteristics of rail transport. For railway identification specifically, we employed a hybrid approach combining numerical analysis and GIS methods. As shown in Figure 2 we extracted railway networks from geospatial data and calculated the proximity of GPS trajectory points to these networks. To improve classification accuracy, we created a 100-meter buffer around railway lines to account for GPS measurement accuracy, and calculated the average distance from each trajectory point with *move* status to the nearest railway segment. Additionally, a trajectory is classified as a railway commuter only if all trajectory points in moving status have their nearest railway line within 5 candidate lines, which prevents misclassification in areas where road and rail networks overlap densely.

This multi-modal classification approach enables detailed analysis of transportation behavior across different user segments. Figure 3 illustrates the aggregated trajectory linestrings for each identified transportation mode. Railway users predominantly travel along railway lines, cars and buses follow more complex road systems, while bicycles and pedestrians exhibit notably shorter path lengths. We validated our classification results by examining Figure 3 and comparing the transportation mode distribution (Supplementary Figure S1) with Japan’s national transportation mode choice report [45]. Our data shows a higher proportion of railway users, likely reflecting sampling bias in smartphone data that potentially excludes individuals under 13 and over 65 years of age, and non-commuting users. After implementing this classification framework across our dataset, we validated our approach through comparison with official statistics.

3.3. Transfer Identification and Validation

The route choice model requires two quantities from each railway commuter’s trajectory: the number of burdensome transfers and the sequence of transfer stations defining each route alternative. We extract both by first reducing the high-resolution status sequence into a compact representation capturing only structural stops, the *commuting motif*, following network-science conventions [46–48], and then identifying which stops correspond to transfers.

The motif representation serves two analytical purposes beyond providing model input. It validates the transfer-identification pipeline at this stage of the refinement chain. We compare the observed $n \geq 3$ share (approximately 10%) against the independent estimate for Tokyo reported by [49], which uses a survey-based transfer definition that, like the motif-level count here, includes both burdensome and strategic (express-switching) transfers. Matching the survey’s definition is what makes the comparison meaningful; the model-specific refinement to burdensome transfers is applied later, in Section 3.4.

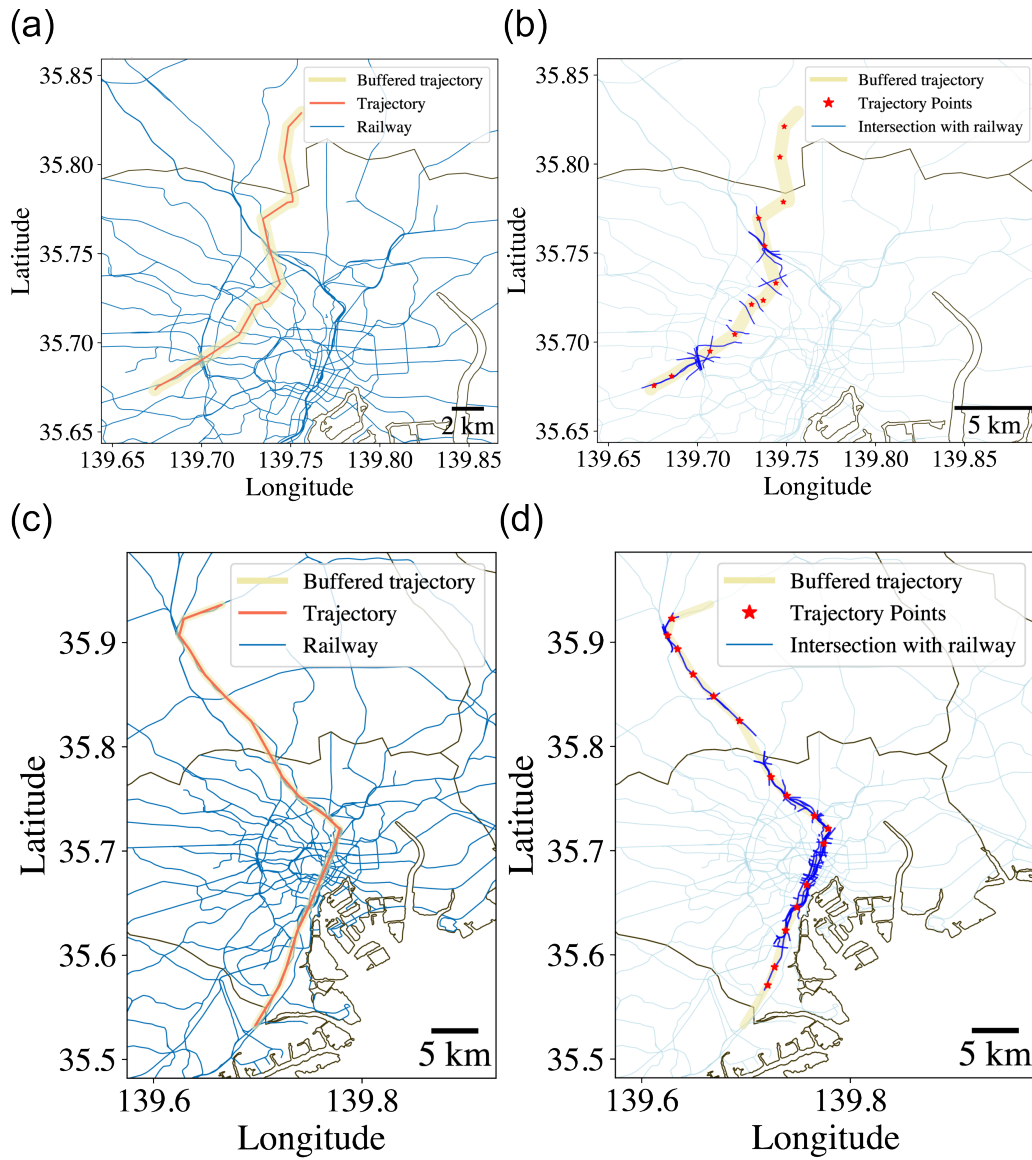


Figure 2. Transportation mode identification process for railway and car commuters. Raw GPS trajectories (orange lines) are overlaid on the railway network (light blue lines) with 100-meter buffered zones (yellow). Dark blue segments indicate intersections between buffered zones and railway lines. Red stars mark identified trajectory points within these intersection areas. **(a,b)** Car commuter example: large distances between *move* status GPS points and railway network result in car mode classification. **(c,d)** Railway commuter example: short distances between *move* status GPS points and railway network result in railway mode classification.

The extraction proceeds in two steps:

- (1) Status sequence simplification: Converting the complete status sequence $\{S(t_i)\}_{i=1}^M$ (with M time points, Section 3.1) into a simplified sequence $\{S'(t_c)\}_{c=1}^m$ (with $m < M$ time points) by retaining only status changes. Each consecutive status must differ from the previous one: $S'(t_c) \neq S'(t_{c-1})$ for all $c > 1$. For example, if the original sequence is (*home–home–home–move–move–work*), the simplified sequence becomes (*home–move–work*).
- (2) Interruption identification: Detecting interruptions where $S'(t_k) = \text{stroll}$ occurs between two movement periods:

$$S'(t_{k-1}) = \text{move} \quad \text{and} \quad S'(t_{k+1}) = \text{move} \quad (4)$$

Through our analysis, we identified several characteristic commuting motifs that represent different degrees of commuting smoothness in Table 4. At this stage, the definitions apply to all transportation modes; for railway commuters specifically, we impose additional spatial constraints below (and summarize the full refinement pipeline in Table 5).

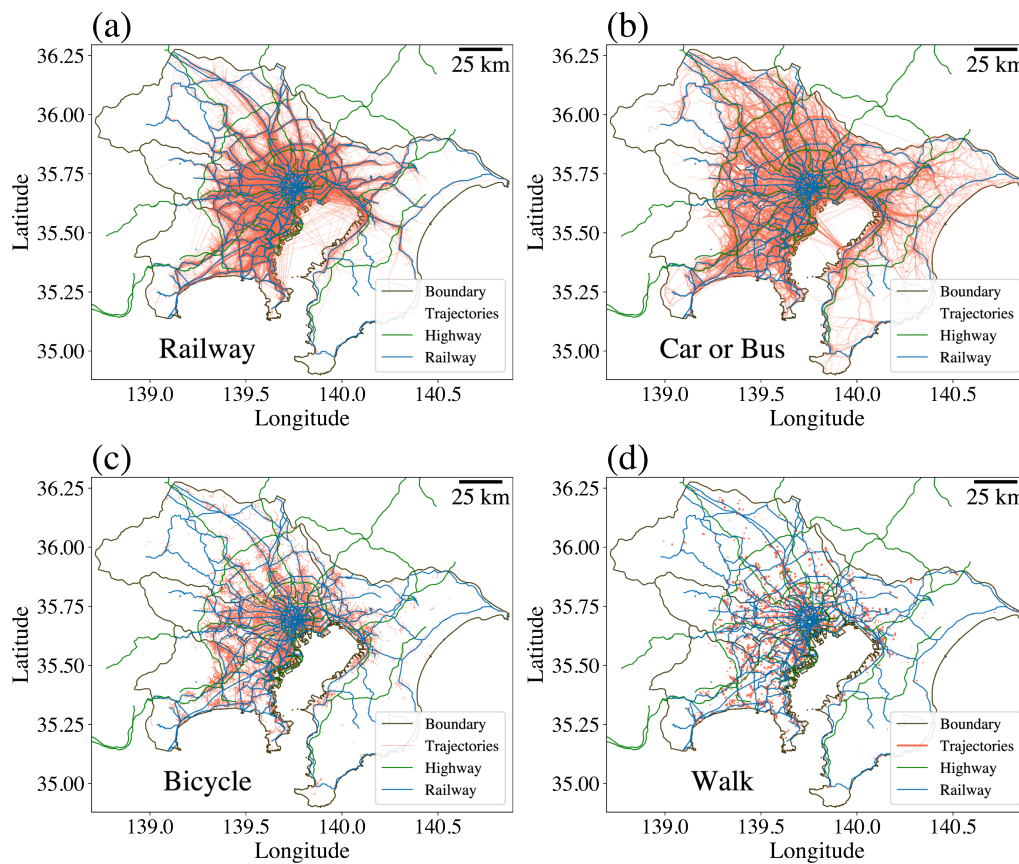


Figure 3. Aggregated trajectories for four recognized transportation modes. (a) Railway trajectories (red lines) closely follow the railway network (blue lines), demonstrating high spatial correlation with rail infrastructure. While some trajectories, such as those across the sea (Tokyo Bay), may appear discontinuous due to signal loss from tunnels or urban canyons, these instances are infrequent and do not compromise the overall dataset integrity for aggregate analysis. (b) Car or bus trajectories closely align the highway network (green lines) and urban road systems, instead of the railway network. Notably, the trajectories across Tokyo Bay align with the shape of the Tokyo Bay Aqua-Line. (c) Bicycle trajectories exhibit intermediate-range mobility patterns with moderate spatial coverage. (d) Walking trajectories display short-distance, localized movement patterns concentrated in urban areas. The distinct spatial characteristics of each mode validate the effectiveness of the transportation mode classification algorithm in this study.

Table 4. Characteristic commuting motifs that represent different degrees of commuting smoothness.

Stop Count	Motif	Description	Route
$n = 0$		Smooth	Home→move→work
$n = 1$		Interrupt once	Home→move→stroll→move→work
$n = 2$		Interrupt twice	Home→move→stroll→move→stroll→move→work
$n \geq 3$		Interrupt multiple	Home → ... → work
Other		Visited over 30 min	Home→stay→work

A pattern *home-stay-work* indicating a prolonged stationary period between home and work locations. These behaviors do not correspond to typical transfer behavior and are excluded from the route-choice model; their inclusion in Table 4 is for pattern display only.

For railway commuters, we additionally require that the *stroll* points fall within a 1 km grid cell containing a railway station, ensuring that only station-area pauses, predominantly transfers are counted.

The 4-min lower bound on the *stroll* duration (Table 2) is chosen to match the average transfer time in Tokyo reported by the MLIT transfer survey [36]: 4.3 min at peak and 3.4 min off-peak. This threshold captures

the majority of genuine transfers while excluding brief operational pauses such as signal adjustments or short platform holds. The 30-min upper bound further excludes non-transfer deviations such as dropping off children or mid-commute meetings, which are assigned to the *stay* status rather than counted as interruptions.

Table 5. Progressive refinement of the interruption concept through the analytical pipeline.

Stage	Section	Definition and Scope
Stroll status	Section 3.1	All modes: Speed < 8 km/h; duration from 4 to 30 min in a 1 km grid cell.
Interruption	Section 3.3	All modes: A stroll between two move states (Equation (4)).
Transfer	Section 3.3	Railway users: An interruption whose stroll points fall within a station-area grid cell ¹ .
Express transfer	Section 3.4	Railway users: A transfer where the transfer route is faster than the fastest direct route for the same OD.
Model variable $NT_{n,j}$	Section 4.2	Railway users only: Number of pure burdensome transfers, excluding express transfers (Equation (10)).

¹ A 1 km grid cell containing a railway station; threshold based on MLIT transfer survey [36].

The distribution in Figure 4 confirms that transfer behavior is the dominant source of route variation in Tokyo's railway network, justifying our focus on transfer counts as a primary attribute in the route-choice model. As a validation, Ref. [49] reports that commutes involving three or more transfers represent approximately 10% of the Tokyo commuting population, closely matching the 10% observed in our $n \geq 3$ bin. The corresponding distribution for non-railway commuters, together with the cross-mode comparison, is provided in Supplementary Figure S13.

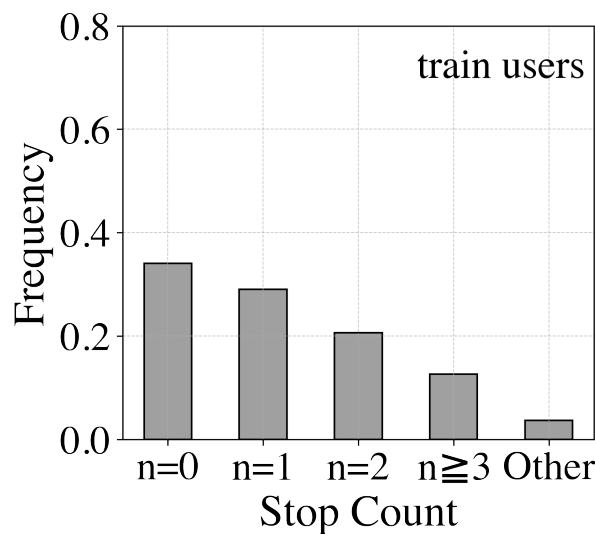


Figure 4. Commuting motifs distribution for railway commuters. Motifs are classified by the number of interruptions (stop count) between home and work, where each interruption occurring within a 1 km grid cell containing a railway station is identified as a transfer. Direct routes ($n = 0$) account for approximately 35%, with single-transfer ($n = 1$) and double-transfer ($n = 2$) routes comprising a substantial share, reflecting the structural complexity of Tokyo's railway network. The *Other* category captures trips containing a *stay* state (pause exceeding 30 min); these are excluded from the route-choice model.

3.4. Route Alternatives and Choice Set Construction

The route-choice analysis is restricted to railway commuters. Each choice situation is defined by a fixed Origin–Destination (OD) pair, spatially anchored to 1 km grid cells containing the commuter's home and work locations. We use 1 km grids rather than station-based Voronoi meshes [7] because the high station density in Tokyo's Central Business District (CBD) produces extremely small, irregular Voronoi cells. These meshes are sensitive to minor GPS errors and frequently misclassify origins and destinations; a visual comparison is provided in Figure 5a. The 1 km grid provides a robust, standardized spatial unit that mitigates GPS noise and guarantees sufficient sample sizes per OD pair.

We distinguish two concepts used throughout the remainder of the paper. A *route alternative* is an abstraction: the commuting strategy shared by all realized commutes within the same OD pair that share the same departure time window and transfer-station sequence. We also refer to the *physical route* as the transfer-station sequence alone, independent of peak-hour status; a single physical route may correspond to two route alternatives (one peak, one off-peak) if it is taken by commuters at both times of day. Travel time and distance of a route alternative are defined as the median over the realized commutes that instantiate it. The choice set faced by each commuter consists of the route alternatives available within their OD pair, and the route-choice model describes which alternative each commuter's trajectory is assumed to represent.

3.4.1. Per-Trajectory Extraction

For each raw trajectory we record:

- (1) Origin and destination. The 1 km grid cells containing the first and last GPS points, identified via the *home* and *work* statuses from Section 3.1.
- (2) Departure time and peak-hour indicator. The timestamp of the first non-*home* point after morning departure. The trajectory is tagged *peak* if this timestamp falls between 07:00 and 10:00, and *off-peak* otherwise.
- (3) Transfer count and transfer stations. The number of transfers (defined in Section 3.3) and the station at which each occurred, encoded by its Voronoi mesh identifier (Figure 5a).

3.4.2. Route Tagging and Aggregation

Each trajectory is assigned a *route tag* combining its peak-hour status, transfer pattern, and transfer-station identifiers. Two trajectories within the same OD pair and with identical tags represent the same route alternative; their travel times and distances are aggregated by the median. Table 6 shows the resulting records for a representative OD pair.

Table 6. Sample of route classification results and feature encoding

OD	Route Tag	Time	Distance	Departure	No. of Transfer	...	Station
OD ₁	peak_direct-1	65.0	60.03	7:10	0	...	—
OD ₁	peak_direct_Vor _{station}	55.0	53.85	10:10	0	...	Vor _{station}
OD ₁	offpeak_transfer_Vor _{station}	59.0	54.85	10:10	1	...	Vor _{station}

Note. The route alternatives for each origin–destination pair together form the choice set. Each row represents a route alternative; its Time and Distance values are the medians over all realized commutes sharing the alternative's tag. The Transfer Station column refers to the Voronoi mesh ID Vor_{station} of the transfer location; when two transfers occur, two separate columns are used. The Voronoi meshes are visualized in Figure 5a. For OD₁, the first row is a genuine non-transfer commute (Transfer Station = −1), while the second row captures a local-to-express switch at Vor_{station}. Although express switching does not increment No. of Transfer, the recorded transfer station produces a distinct route tag, preserving the slow-train vs. express-train distinction.

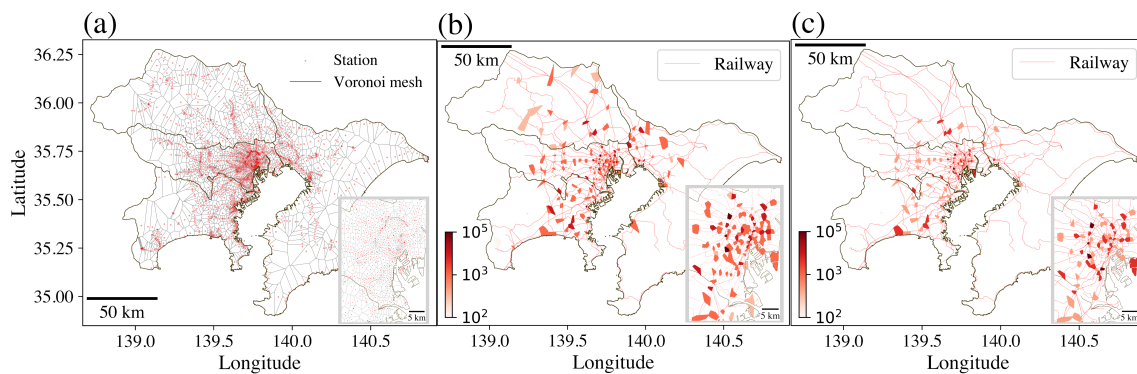


Figure 5. Voronoi tessellation of railway stations and geographical distribution of transfer stations in the study area. (a) Voronoi cells constructed based on station locations, with red points indicating the nearest station for each cell. (b) Geographical distribution of first transfer stations; the inset shows the zoomed Tokyo Central Business District (CBD). The color scale represents the number of commuters making their first (or only) transfer at each station. (c) Geographical distribution of second transfer stations. High-usage stations are more concentrated toward the CBD compared to first transfer stations.

Two filters are applied before estimation. First, OD pairs with fewer than 20 realized commutes are excluded, and the sensitivity test to test whether 20 is statistically sufficient for the model is provided in Supplementary Section 10 (Robustness Check). Second, OD pairs with fewer than 2 or more than 6 physical routes are excluded, following [50]: single-route OD pairs contain no choice to model, while OD pairs with very large choice sets typically reflect non-routine travel or GPS noise rather than stable commuting behavior.

3.4.3. Express-Transition Adjustment

Tokyo's railway network runs local and express services on many lines, so a small number of transfers are strategic rather than burdensome: a commuter accepts a transfer in order to board a faster express train. Because such transfers save time rather than add cost, we reclassify them before finalizing the transfer count:

- (1) For each OD pair, find the travel time of the fastest route alternative whose tag indicates no transfer.
- (2) Any route alternative with one or more transfers but a median travel time lower than this benchmark is reclassified: its transfer count is set to zero and its transfer-status tag component is reset to *direct*, while the transfer-station identifiers are retained. The stations are kept in the tag to preserve the distinction between local and express services sharing the same endpoints.
- (3) After this adjustment, all transfers remaining in the dataset are burdensome, i.e. time-adding.

As illustrated in the second row of Table 6, an express transition produces a record with zero transfers but a non-empty transfer-station field.

Because transfers are identified through successive filtering stages, we summarize the progressive refinement in Table 5 to clarify how each stage narrows the concept.

4. Model

Our modeling framework employs a discrete choice approach based on the Multinomial Logit (MNL) model [14, 15], utilizing the route and feature information constructed in Section 3. Commuters are assumed to select routes that minimize a generalized cost incorporating travel time, transfers, and crowding. Section 4.1 introduces the MNL framework with an explicit scale-direction decomposition of the estimated parameter vector. Section 4.2 specifies the route attributes entering the model and the rationale for their selection. Section 4.3 describes parameter estimation, the two estimation schemes (Model 1 and Model 2), and the metrics used for model evaluation.

4.1. Route Choice Model

We model route choice using a Multinomial Logit (MNL) framework [14, 15] in which commuters select routes that minimize a generalized energy cost. Let C_n denote the set of feasible routes available to commuter n . The cost of route $j \in C_n$ is modeled as a linear combination of its attributes:

$$E_{nj} = \mathbf{X}_{nj}^\top \boldsymbol{\omega} \quad (5)$$

where \mathbf{X}_{nj} and $\boldsymbol{\omega}$ are vectors of the same dimension, determined by the number of route attributes. \mathbf{X}_{nj} contains the attribute values of route j for commuter n (e.g., travel time, number of transfers), and $\boldsymbol{\omega}$ is the weight vector quantifying each attribute's contribution to the total cost, a normalized direction vector. A positive element of $\boldsymbol{\omega}$ indicates that the corresponding attribute increases cost and thus reduces choice probability; a negative element indicates the opposite.

The probability that commuter n chooses route j follows the MNL form:

$$P_{nj} = \frac{\exp(-\beta E_{nj})}{\sum_{k \in C_n} \exp(-\beta E_{nk})} \quad (6)$$

Equation (6) can be derived from the standard random utility framework [15, 16], which makes the role of β explicit. Each commuter n is assumed to associate each route j with a utility

$$U_{nj} = -E_{nj} + \varepsilon_{nj} \quad (7)$$

where E_{nj} is the deterministic cost defined in Equation (5) and ε_{nj} is an unobserved error term capturing idiosyncratic factors. When ε_{nj} is i.i.d. and Gumbel-distributed with scale parameter μ , integrating over the error yields:

$$P_{nj} = \frac{\exp(-E_{nj}/\mu)}{\sum_{k \in C_n} \exp(-E_{nk}/\mu)} \quad (8)$$

which recovers Equation (6) with $\beta = 1/\mu$. In this view, β is the inverse of the Gumbel error scale, controlling how sharply the deterministic cost dominates the unobserved stochastic component.

In standard applications, the error scale is typically absorbed into the utility coefficients, preventing separate identification from choice data [15]. Rather than claiming structural identification, we apply a standardization convention by constraining the ℓ_2 -norm of the weight vector ($\|\boldsymbol{\omega}\|_2 = 1$). In practice, after estimating an unconstrained vector $\boldsymbol{\omega}_{\text{raw}}$ via maximum likelihood, we define:

$$\beta = \|\omega_{\text{raw}}\|_2, \quad \omega = \frac{\omega_{\text{raw}}}{\|\omega_{\text{raw}}\|_2} \quad (9)$$

Fixing $\|\omega\|_2 = 1$ is strictly an identifying convention. It places β on a standardized scale, enabling direct comparisons of *choice determinism* across OD pairs, months, and studies that adopt the same $[0, 1]$ feature scaling. Under this shared metric, a larger β indicates that collective choices respond more sharply to cost differences, while a smaller β approaches random selection.

Under this convention, β inherits the interpretation from Equation (8) as the inverse Gumbel error scale: it measures how sharply the population's collective choices respond to cost differences, with larger β producing more deterministic choices and smaller β approaching random selection. Meanwhile, ω captures the relative importance of each route attribute.

The MNL specification in Equation (6) carries the Independence of Irrelevant Alternatives (IIA) assumption: the relative odds of any two alternatives depend only on those two alternatives' attributes, not on the presence or attributes of a third [15]. IIA becomes problematic when alternatives share substantial unobserved features, such as overlapping physical segments in a route-choice context. Two modeling choices in this study mitigate the assumption's consequences. First, we define alternatives at the category level (peak status, transfer-station sequence), which aggregates partially overlapping physical paths into structurally distinct options and reduces link-level redundancy. Second, the OD-specific estimation variant (Section 4.3.1) absorbs correlation effects that are local to each network topology, which a pooled model would have to treat as IIA violations. A full path-size correction [51] remains a direction for future work (Section 6).

The equivalence between this MNL formulation and the canonical ensemble [30,52] of statistical physics is discussed in Section 6; the full Maximum Entropy derivation is provided in Appendix B.

4.2. Feature Vector Specification

4.2.1. Variables Included

For commuter n choosing route j , the deterministic attribute vector is

$$\mathbf{X}_{nj} = (\text{CT}_{nj}, \text{CD}_{nj}, \text{PK}_{nj}, \text{NT}_{nj}, \text{CAP}_{nj1}, \text{CAP}_{nj2}), \quad (10)$$

with each variable defined as follows:

- CT_{nj} : door-to-door commuting time (minutes).
- CD_{nj} : commuting distance (kilometers).
- PK_{nj} : peak-hour indicator (1 if departure is between 07:00–10:00, 0 otherwise).
- NT_{nj} : number of burdensome transfers, as refined in Table 5.
- $\text{CAP}_{nj1}, \text{CAP}_{nj2}$: capacity at the first and second transfer stations (defined below).

These six variables capture the two dimensions commuters trade off in route choice: *efficiency* (time, distance) and *inconvenience* (transfer count, peak-hour demand pressure, and station capacity as a proxy for crowding). The remainder of this subsection gives the rationale for each variable, the candidate variables we considered and excluded, and the final specification selected by empirical testing.

4.2.2. Why These Variables

Travel time CT and distance CD are standard components of generalized travel cost in transportation research [15]. We retain both rather than collapsing them into a single composite because time and distance separately carry the non-monetary and monetary dimensions of route disutility in the Tokyo context: fares in Japan are largely distance-determined [53], so CD effectively proxies the monetary cost component without requiring fare data at the route level.

Transfer count NT enters because transfers impose a well-documented psychological and physical penalty beyond the added travel time [18]. Our refined definition (Table 5) counts only burdensome transfers, ensuring NT reflects the kind of transfer behavior the model is meant to explain.

Peak-hour status PK is included because departure time correlates systematically with crowding, service frequency, and headway variability, all of which affect perceived route disutility independently of the other variables.

Station capacity CAP_k addresses the role of crowding in route choice while avoiding the simultaneity bias inherent in real-time crowding measures. Because observed crowding is itself the outcome of collective route choices, using real-time counts would make the explanatory variable endogenous to the choice being modeled.

Instead we define station capacity as the maximum hourly population observed at station s during the study year:

$$CAP_s = \max_{t \in \text{study year 2023}} Pop_{st}, \quad (11)$$

where Pop_{st} is the hourly passenger count at station s during hour t . This measure is predetermined with respect to any individual commuter's choice and therefore exogenous, while still capturing the infrastructural crowding pressure commuters experience at major hubs. We enter CAP_1 and CAP_2 as separate variables rather than a single composite because first and second transfer stations play asymmetric roles in Tokyo's hierarchical network (see Section 5.1.3).

A note on congestion. Because Tokyo's railway network operates on grade-separated infrastructure independent of the road system, road-traffic congestion does not enter the model: railway travel time and route choice are decoupled from road conditions. Congestion in this paper therefore refers to station-level density and transfer complexity, not roadway flow.

4.2.3. Variables Considered and Excluded

Two further candidate variables were considered and excluded:

Monetary cost at the route level. For a fixed OD pair, fares are typically uniform across lines operated by the same company, and most employers reimburse commuting passes (*tsūkin teikiken*), so out-of-pocket cost exerts limited influence on daily decisions. The remaining distance-driven variation in cost is already captured by CD.

Higher-order transfer capacity (CAP_3 , CAP_4 , etc.). Only about 10% of observed commutes involve more than two transfers. Industry survey data [49] confirms that regular commutes with three or more transfers are rare; properties requiring such commutes are generally recommended only for full-time remote workers. Most of the 10% in our data likely represent non-routine trips (e.g., daytime client meetings) rather than regular commuting, so adding CAP_3 or CAP_4 would fit noise rather than structure.

4.2.4. Final Specification

Beyond rejecting these individual candidates, we systematically compared ten alternative functional specifications of the retained variables, including linear-vs-logarithmic transformations of CD and CAP, aggregating CAP_1 and CAP_2 into a single maximum, and including higher-order capacity terms. Details are provided in Supplementary Table S2.

The specification test selects logarithmic transformations of distance and capacity, yielding the estimated form

$$\mathbf{X}_{nj} = (\text{CT}_{nj}, \log \text{CD}_{nj}, \text{PK}_{nj}, \text{NT}_{nj}, \log \text{CAP}_{nj1}, \log \text{CAP}_{nj2}), \quad (12)$$

Multicollinearity diagnostics confirm that the retained variables are non-redundant (all variance inflation factors below 5, the conventional threshold [54,55]; see Supplementary Table S1).

4.3. Model Estimation and Evaluation

We estimate model parameters by maximizing the following sample log-likelihood

$$\ell(\boldsymbol{\omega}_{\text{raw}}) = \sum_{n=1}^N \sum_{j \in C_n} y_{nj} \log P_{nj} \quad (13)$$

where $y_{nj} = 1$ if commuter n chooses route j and 0 otherwise. The optimization is performed on the raw, unconstrained weight vector $\boldsymbol{\omega}_{\text{raw}}$ without any direct constraints. Then we obtain β and $\boldsymbol{\omega}$ by Equation (9).

We compute the standard error of the estimated parameter $\boldsymbol{\omega}_{\text{raw}}$ using the inverse of the Hessian matrix of the log-likelihood according to the asymptotic theory of the maximum likelihood estimator. Then we use the delta method [56,57] to compute the standard error of the normalized parameters β and $\boldsymbol{\omega}$ (see Appendix for detail procedure).

4.3.1. Two Estimation Schemes

We estimate the model in two variants:

- Common Parameter Over all ODs (Model 1). A single parameter vector $\boldsymbol{\omega}_{\text{raw}}$ estimated using all routes pooled across all OD pairs. One estimate per month.
- OD-Specific Parameters (Model 2). A separate parameter vector $(\boldsymbol{\omega}_{\text{raw}})_{\text{OD}}$ estimated for each OD pair. One estimate per OD pair.

The two variants share the same mathematical form (Equation (6)) but are estimated at different scopes, which

lets them answer different questions. Model 1 measures whether system-wide collective determinism is stable over time: by pooling all OD pairs each month and comparing the resulting twelve monthly estimates, it speaks to the temporal dimension of our research question (Section 5.1). Model 2 measures whether determinism varies with commuting context: by estimating parameters per OD pair, it exposes systematic heterogeneity that a pooled model absorbs into its unobserved random term (Section 5.2.1, Section 5.2.2). Neither variant subsumes the other—they estimate the same object at different granularities, and the Model 1 estimate serves as the aggregate benchmark against which the mode of the Model 2 distribution is validated.

Normalization and the β Baseline

Prior to estimation, each route attribute is Min–Max normalized to the interval $[0, 1]$. For Model 1, the minimum and maximum are computed across all routes in the study area; for Model 2, within each OD pair. Both variants therefore share the same $[0, 1]$ scale, which gives β a concrete interpretation: it measures choice sensitivity when the worst-to-best cost gap is unity. This normalization convention allows direct comparison of β estimates across OD pairs within this study, across months, and across different metropolitan networks that adopt the same convention. The city-wide Model 1 estimate β_{city} thus serves as an empirical baseline against which each OD-specific β^{OD} can be judged (Section 5.2): $\beta^{\text{OD}} > \beta_{\text{city}}$ indicates that a given corridor’s commuters make more deterministic choices than the metropolitan average, while $\beta^{\text{OD}} < \beta_{\text{city}}$ indicates greater stochasticity.

4.3.2. Model Evaluation

To assess the performance of our estimated model, we employ two complementary metrics:

- (1) McFadden’s pseudo- R^2 , calculated as

$$R_{\text{McF}}^2 = 1 - \frac{\ell_{\text{full}}}{\ell_{\text{null}}}$$

where ℓ_{full} is the log-likelihood of the estimated model and ℓ_{null} is the log-likelihood of a null model with equal alternative probabilities [14].

- (2) Hit rate, measuring the proportion of correctly predicted choices:

$$\text{Hit Rate} = \frac{\text{Number of Correct Predictions}}{\text{Number of Choice Situations}}$$

This metric directly reflects the model’s ability to reproduce individual choices [15–17].

Together, these metrics provide a comprehensive evaluation of how well our model captures railway commuters’ path choice behavior in the complex Tokyo transit network.

5. Results

This section presents the results of the railway commuting route choice model, revealing that Tokyo commuters exhibit highly structured, yet fundamentally non-deterministic, preferences that remains stable throughout the year. The following subsections establish this result and then examine its variation across OD pairs. We begin with an aggregated-level model that pools all Origin-Destination (OD) pairs (Section 5.1). This initial analysis addresses three key questions regarding overall commuter behavior: (1) Overall determinism: How deterministic are commuters in the aggregate? (Section 5.1.1); (2) Temporal Consistency: How consistent is overall choice behavior over time? (Section 5.1.2); (3) Attribute Preference: Which route attributes are regarded attractive by commuters? (Section 5.1.3) with a case analysis (Section 5.1.4).

Following the aggregate analysis, we apply the same model structure to individual OD pairs to examine heterogeneity and address: (1) Existence of Heterogeneity: Can OD pairs be grouped by similar choice patterns? (Section 5.2); (2) Behavioral Drivers of Heterogeneity: Which OD pair characteristics account for different choice behaviors, including the time-equivalent cost of transfers across specific OD pairs? (Sections 5.2.1, 5.2.2)

Due to computational constraints in analyzing the massive 12-month trajectory dataset, a representative month was selected for most analyses. Except for the temporal consistency analysis (Section 5.1.2), which uses all 12 months to examine stability over time, all other results are based on April 2023 data. April was chosen as the most representative month due to its largest number of observed users, maximizing statistical reliability of OD-specific estimates. The complete set of estimated parameters (same format as Table 7) for all 12 months is provided in Supplementary Section 4, confirming month-to-month stability of core findings.

Table 7. Estimation results of railway commuting route 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 ($\log(CD_{nj})$, km)	0.579 ***	0.034	133.157
Peak Departure (PK_{nj} , dummy)	-0.117 ***	0.005	-209.778
Capacity at 1st transfer ($\log(CAP_{nj1})$)	0.523 *	0.017	2.543
Capacity at 2nd transfer ($\log(CAP_{nj2})$)	-0.340 ***	0.013	-210.615
Number of users		125,368	
Number of routes		23,887	
Number of ODs		5,908	
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 R^2		0.232	
Hit rate		55.8%	
Spearman ρ		0.724	

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

5.1. Common Parameters for All ODs: Model Validation and Performance

As mentioned in Section 4.3.1, the aggregate model demonstrates strong predictive performance and statistical significance (Table 7), with reliability established through multiple validation metrics.

All explanatory variables are statistically significant according to their t-statistics and p-values (Table 7). The model correctly predicts the chosen route in 55.8% of observations, substantially exceeding the 16.7% random benchmark (based on a max of six alternatives per OD pair). The McFadden's pseudo R^2 of 0.232 is acceptable for empirical social science modeling [58].

Most importantly for a discrete choice model, the model achieves a high Spearman correlation coefficient (ρ) of 0.72 between predicted and observed route choice probabilities (Figure 6b). This metric confirms that the model accurately captures the relative ranking of alternatives, assigning higher probabilities to routes that are chosen more frequently, the fundamental objective in understanding choice behavior.

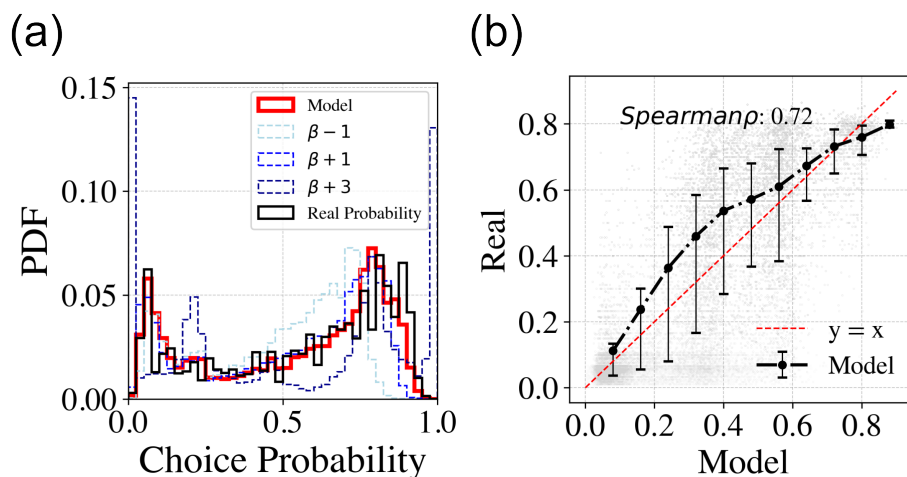


Figure 6. Model validation and sensitivity to scale parameter β . (a) Predicted route choice probability distribution and its relationship with scale parameter β . The red line shows model predictions with the estimated β value; the light blue line shows the scenario with β decreased by 1; darker blue lines show β increased by 1 and 3, respectively. The dashed black line represents the observed probability distribution. (b) Model validation showing predicted versus observed route choice probabilities. Black dots represent aggregated predictions with error bars indicating the 25th and 75th percentiles; grey dots show individual observations. The red dashed line indicates perfect prediction ($y = x$). Spearman $\rho = 0.72$.

The moderate McFadden’s pseudo R^2 can be attributed to three factors: (1) subtle differences between route alternatives in Tokyo’s dense railway network; (2) spatial aggregation causing within-group heterogeneity in route variables; and (3) inherent limitations in GPS data precision.

5.1.1. Commuter Choice Determinism

Under the standard Random Utility Maximization framework, β is proportional to the inverse of the Gumbel scale parameter, quantifying the variance of unobserved idiosyncratic factors (ϵ_{nj} in Equation (7)) relative to deterministic route costs. Behaviorally, β measures *choice determinism*: the extent to which the commuter population relies on observed cost attributes rather than unmeasured preferences or habits.

High β : The system is highly deterministic. Even small cost differences produce large probability differences, meaning commuters almost always select the lowest-cost route.

Low β : The system is highly stochastic (random). Cost differences have minimal impact on choices, meaning commuters are nearly equally likely to select high-cost and low-cost routes.

To make the interpretive scale of β concrete, we examine a standardized scenario: a two-route OD pair where the alternatives differ by exactly one generalized energy unit ($\Delta E = 1$). Under the MNL form (Equation (6)), the odds ratio of choosing the better alternative over the worse one ($P_{\text{better}}/P_{\text{worse}}$) is exactly $e^{-\beta(-\Delta E)} = e^\beta$, meaning the better route is e^β times more likely to be chosen. Consequently, since there are only two routes, the probability of selecting the better route evaluates to $P_{\text{better}} = e^\beta/(1 + e^\beta) = 1/(1 + e^{-\beta})$. Evaluating this probability at the empirical Tokyo value, and at its ± 1 and $+3$ perturbations visualized in Figure 6a, yields the following correspondence between β and collective route choice probability (Table 8):

Table 8. The correspondence between β and choice preference.

β	e^β	P_{Better}	Interpretation
0	1	50%	Fully random (all routes equally weighted)
0.658 ($\beta - 1$)	1.93	$\approx 66\%$	Near-random; modest preference for better route
1.658	5.25	$\approx 84\%$	This study (Tokyo, city-wide)
2.658 ($\beta + 1$)	14.3	$\approx 93\%$	Substantially more deterministic
4.658 ($\beta + 3$)	105	$\approx 99\%$	Near-deterministic
$\rightarrow \infty$	$\rightarrow \infty$	$\rightarrow 100\%$	Perfectly deterministic

Note: Bold row indicates the empirical estimate from this study (Tokyo, city-wide, $\beta = 1.658$). At the Tokyo empirical estimate, a 1-unit advantage in generalized energy means that the better route is 5.25 times more likely to be chosen than its alternative.

Therefore, β serves as an aggregate measure of choice determinism, the degree to which commuters consistently select routes in alignment with the calculated cost function.

As shown in Figure 6a, both the model predictions (red line) and observed probabilities (black line) exhibit a distinctive two-peak pattern: a smaller peak near zero probability and a larger peak near one. This bimodal distribution reveals the underlying choice structure, as most routes either have low choice probabilities (unlikely alternatives) or relatively high probabilities (preferred alternatives), with few routes falling in the intermediate range. This pattern indicates that commuters make relatively clear distinctions between desirable and undesirable routes based on their attributes, rather than distributing choices uniformly across alternatives. At the same time, the peak near probability 1 does not saturate at 1, and the peak near 0 does not collapse to 0, a non-negligible mass remains across the full probability range. Tokyo commuters thus exhibit a strong but non-deterministic preference: clearly biased toward low-cost routes, yet far from the strict cost minimization that a purely deterministic system would produce.

The close alignment between the red and black lines demonstrates that the estimated β value (1.658) accurately captures this choice behavior. Figure 6a further illustrates how variations in β affect the probability distribution. As β increases to 3 (darker blue lines), the system becomes near-deterministic: low-cost routes are assigned probabilities approaching 1, while high-cost routes approach zero, resulting in more pronounced peaks at the extremes. Conversely, decreasing β (light blue line) shifts the distribution toward less deterministic choices, with more probability mass concentrated around 0.5, indicating greater randomness in route selection. The observed distribution sits clearly between these two extremes, visually confirming that Tokyo’s choice determinism level is substantial but not absolute.

The next section illustrates if the current β value of 1.658 positions the system at a balanced state or not.

5.1.2. Consistency of Choice Behavior

This section evaluates the temporal consistency of commuters' overall determinism and route choice strategy by examining the monthly stability of the estimated model parameters.

The scale parameter β demonstrates remarkable consistency throughout the year, with monthly values clustering around the annual mean of 1.60 (Figure 7a). This temporal consistency indicates that the aggregate level of choice determinism and the sensitivity to route costs remain relatively constant across seasons.

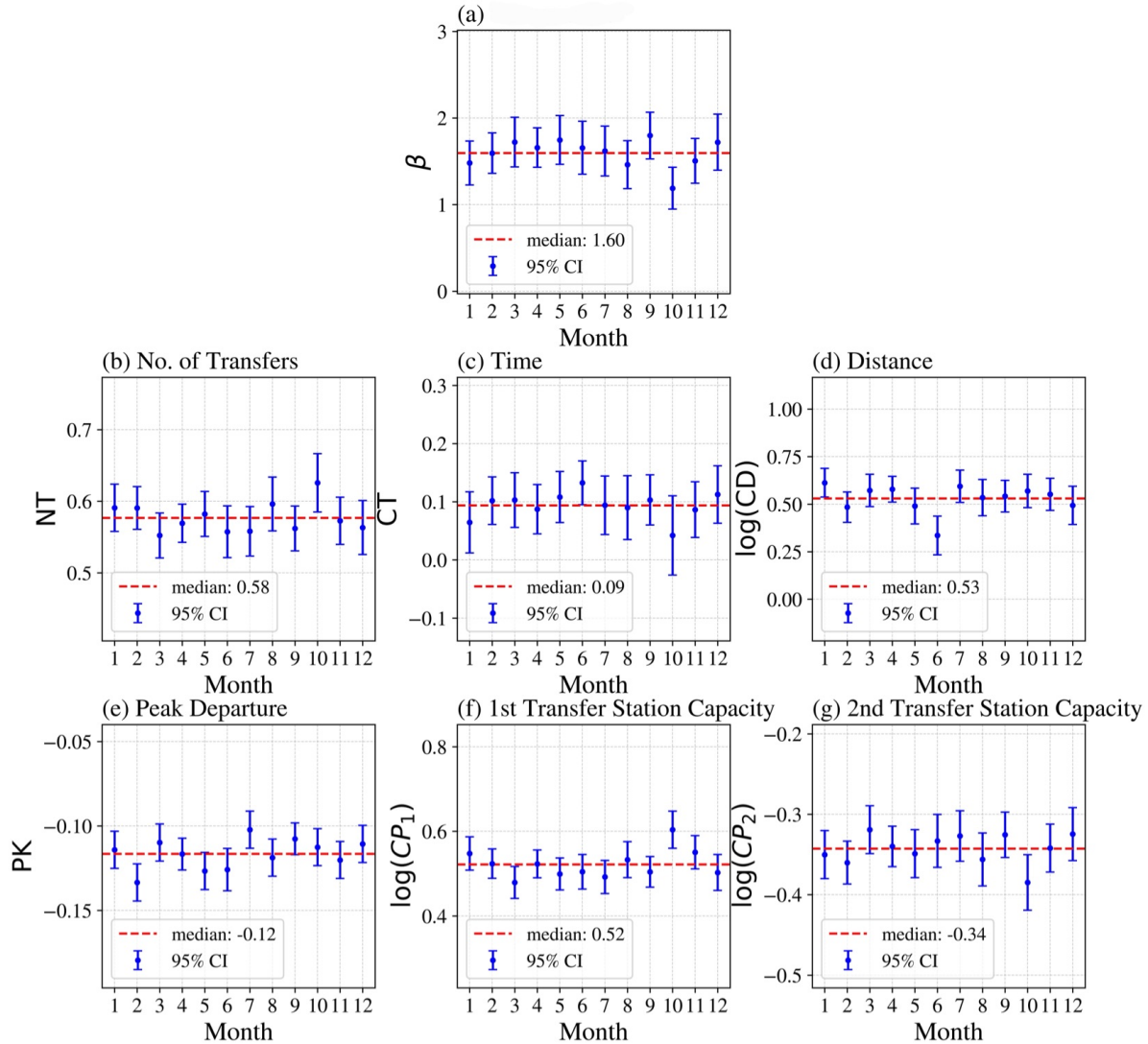


Figure 7. Monthly estimates and 95% confidence intervals (CI) (calculated by delta method) for the normalized weights, ω , across twelve months. (a) Inverse of the unobserved error scale (β) ranges from 1.189 (October) to 1.797 (September) with median 1.60; confidence intervals overlap across all months. (b) Number of transfers coefficient (NT) ranges from 0.552 (March) to 0.626 (October) with median 0.58; confidence intervals generally overlap throughout the year. (c) Time coefficient ranges from 0.042 (October) to 0.133 (June) with median 0.09; confidence intervals overlap across months. (d) Distance coefficient ($\log(CD)$) ranges from 0.336 (June) to 0.613 (January) with median 0.53; confidence intervals overlap for all months. (e) Peak departure coefficient (PK) ranges from -0.134 (February) to -0.102 (July) with median -0.12 ; confidence intervals overlap throughout the study period. (f) First transfer station capacity ($\log(CP_1)$) ranges from 0.479 (March) to 0.604 (October) with median 0.52; confidence intervals overlap across all months. (g) Second transfer station capacity ($\log(CP_2)$) ranges from -0.385 (October) to -0.319 (March) with median -0.34 ; confidence intervals overlap for all months. The dashed red line in each subplot indicates the annual median value.

Similarly, the normalized coefficients (ω) for all other route attributes (Figure 7b–g) exhibit minor fluctuations, confirming the stability of the relative importance of factors like travel time, transfer frequency, and station capacity. The narrow and overlapping 95% confidence intervals (CIs) (See Method Section 4.3) across all twelve months provide strong statistical evidence that the identified behavioral patterns represent stable structural features of

commuter decision-making, rather than being artifacts of specific time periods or data collection conditions.

Despite this overall stability, we observe modest seasonal fluctuations. The lowest β value occurs in October (approximately 1.20), suggesting slightly increased randomness in route choices, likely attributable to Japan's momiji (autumn foliage) season when tourist activity may disrupt normal transportation patterns. Conversely, the highest β value appears in September (approximately 1.80), indicating more deterministic choice behavior, which correlates with the resumption of highly structured commuting routines following the summer period.

To assess the practical significance of these seasonal variations in commuter route choice determinism, Figure 6a examines how changes in β affect the predicted probability distribution. The observed monthly fluctuations (approximately ± 0.3 around the median β value) produce minimal changes in the predicted choice probabilities, demonstrating model robustness. These limited effects indicate that seasonal variations have negligible practical impact on route choice predictions.

This temporal consistency confirms that the identified route choice patterns represent stable structural features of commuter behavior rather than temporary phenomena.

5.1.3. General Route Attribute Preferences

The estimation results, summarized in Table 7, reveal system-wide commuter behavioral preferences and directly identify which route attributes constitute a higher or lower *energy cost* in the choice decision. Consistent with a model seeking to minimize route cost, the signs of the coefficients reveal the general attractiveness or deterrent effect of each attribute.

Route Deterrents (Positive Coefficients)

Three fundamental travel energy costs consistently deter route selection: The positive coefficients for number of transfers (NT_{nj}), commuting time (CT_{nj}), and log commuting distance ($\log(CD_{nj})$) confirm that commuters system-wide prefer routes that minimize transfers, travel time, and distance.

Notably, the positive coefficient for $\log(CAP_{nj1})$ (Capacity at 1st transfer or single transfer) reveals that transferring at high-capacity stations imposes higher perceived costs. This counterintuitive result likely reflects the compounding negative effects experienced at major transfer hubs: difficulty boarding crowded trains, low probability of securing a seat, longer walking distances between platforms, and greater uncertainty during service disruptions. These factors collectively outweigh any potential benefits of higher service frequency at large stations during the initial transfer.

Route Attractions (Negative Coefficients)

Two attributes significantly reduce perceived route costs: The negative coefficient for second transfer station capacity ($\log(CAP_{nj2})$) indicates that high capacity at the second transfer point reduces perceived travel burden. This asymmetric effect compared to first transfers reflects the distinct spatial and functional roles of transfer locations. As shown in Figure 5, second transfers predominantly occur at major CBD hubs (e.g., Shinjuku, Shibuya) where high capacity is structurally coupled with superior service frequency and better infrastructure. This suggests that commuters have adapted their expectations, perceiving high capacity at these specific, centralized locations as an expected feature that facilitates the final leg of the journey, rather than a deterrent.

The negative coefficient for peak departure (PK_{nj}) reflects a structural constraint of commuting demand. The majority of trips are temporally fixed to peak hours due to work schedules, effectively making peak-hour travel a *default* and therefore less costly (or necessary) choice relative to the perceived cost of missing the required arrival time. While peak-hour departure is often constrained by work schedules, the rise of telework and flexible shifting in post-pandemic Tokyo has introduced greater temporal choice, making the PK_{nj} variable a relevant indicator of shifted commuting constraints. The negative and significant coefficient suggests that temporal constraints often override comfort considerations in route selection for the commuters.

Approximately 10% of commuters in our sample make more than two transfers. We did not include third or fourth transfer capacity variables for two reasons. First, the sample size for these cases is too small for reliable statistical estimation. Second, trips with multiple transfers often represent non-routine activities (e.g., client meetings, special errands) rather than regular commuting patterns. Our model focuses on typical commuting behavior to maintain generalizability and interpretability.

5.1.4. Case Analysis

To validate the model's predictive capability and capture nuanced commuter behavior, we analyzed a specific origin-destination (OD) pair: Musashi-Shinjō to Shinjuku. We selected this case based on three criteria: (i) Sufficient

sample size for stable monthly probability estimates (typically 60+ commuters; Table 9); (ii) A choice set of at least four physically distinct alternatives, making the asymmetric roles of transfer station capacity observable within a single case. This OD pair provides a representative lens into the complex trade-offs commuters make between travel time, distance, and transfers.

As shown in Figure 8, the model’s predicted route usage (Figure 8b) closely aligns with the actual route usage derived from GPS data (Figure 8a). The most utilized route, R1, which corresponds to the Odakyu Line, exhibits high predicted usage consistent with its superior attributes, including the shortest travel time and a single transfer. Conversely, the least utilized route, R3 (Tōyoko Line), which involves two transfers and a longer travel time, is correctly assigned a low probability by our model.

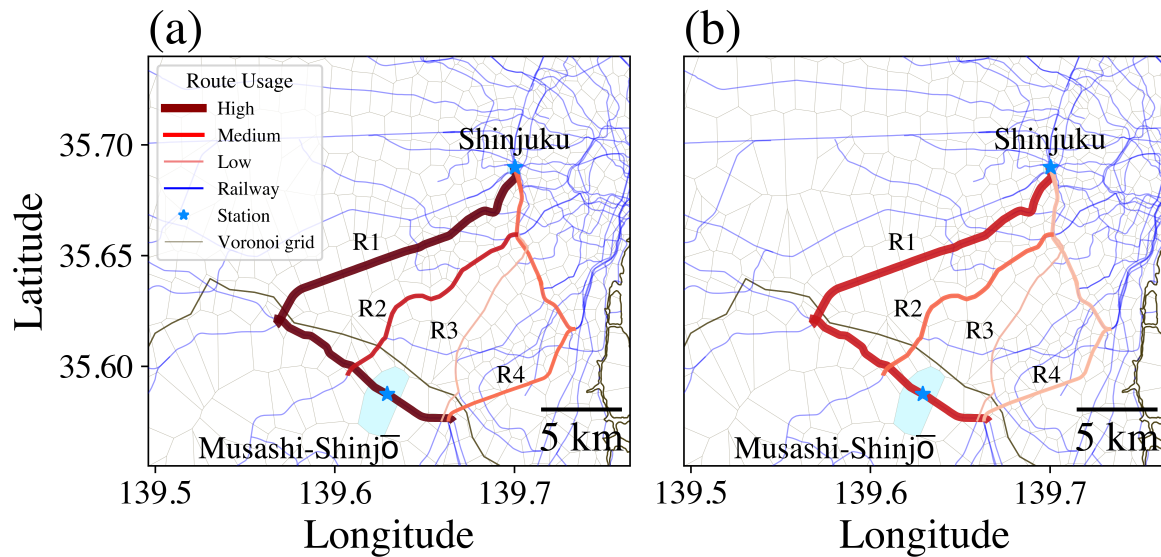


Figure 8. Real and model-predicted route choice maps for the Musashi–Shinjō to Shinjuku OD pair. (a) Actual route usage based on GPS data showing high, medium, and low-usage routes. (b) Route usage predicted by the Model 1 (city-wide pooled estimation) model variants. Line thickness and darkness indicate usage intensity. Four primary route alternatives are examined (detailed in Table 9). The model successfully captures the most and least utilized routes, with predicted patterns closely resembling observed usage.

Table 9. Route options from left to right visualized in Figure 8 for Musashi-Shinjō to Shinjuku OD pair (at 7:00 am).

Route ID	Railway Lines	NT (Times)	CT (min)	CD log(CD) (km)	CAP ₁ log(CAP ₁) (Persons)	CAP ₂ log(CAP ₂) (Persons)
R1	Nambu→ Odakyu	1	33	22 (3.09)	110,466 (11.61)	-
R2	Nambu→ Den-en-toshi→ Yamanote	2	39	19.6 (2.98)	169,774 (12.04)	245,291 (12.41)
R3	Nambu→ Tōyoko→ Yamanote	2	41	18.2 (2.90)	169,898 (12.04)	245,291 (12.41)
R4	Nambu→ Shōnan-Shinjuku	1	34	23.6 (3.16)	169,898 (12.04)	-

Note: Transfer stations and their approximate daily passenger capacity rankings are as follows: (1) Shibuya (2nd transfer for R2, R3); (2) Musashi-Kosugi (1st transfer for R3, R4); (3) Musashi-Mizonokuchi (1st transfer for R2); (4) Noborito (1st transfer for R1).

Table 10 further quantifies these results by showing the monthly observed route choices for the four primary alternatives detailed in Table 9. As visually depicted in Figure 8, R1 is the most frequently chosen option, while R2 and R4 are selected at a roughly equal, moderate rate, and R3 is the least preferred. The *Model Predicted* row, calculated using the route features from Table 9 and the parameter estimates from the model shown in Table 7, successfully captures this relative popularity, correctly identifying the order of preference as $R1 > R2 \simeq R4 > R3$.

A more subtle and insightful observation arises from the intermediate route choices. The Den-en-toshi Line is also commonly used despite an extra transfer, with usage comparable to the Shōnan-Shinjuku Line’s route with a single transfer. This occurs because the Den-en-toshi Line’s transfer happens at Shibuya, a major station with large capacity and numerous accessible lines. Our model’s positive coefficient for $\log(CP_2)$ correctly captures the positive effect of high-capacity second transfer station.

The choice to exclude monetary cost as a feature warrants specific justification. As detailed in Section 3, the widespread use of employer-reimbursed commuting passes in Japan, combined with the typically uniform fare structures across major rail networks, suggests that monetary cost is often not the primary decision variable for the average commuter. Empirical data from specific OD pairs, such as Musashi-Shinjō to Shinjuku, supports this structural assumption. Within this pair, routes R2 (540 JPY) and R3 (560 JPY) are frequently chosen by commuters over the less expensive routes R1 (450 JPY) and R4 (410 JPY). This behavior, where a substantial number of commuters consistently select a route that is approximately 100 JPY more expensive, provides strong revealed preference evidence. This difference in price is clearly outweighed by the non-monetary costs captured in our model, such as travel time, transfers, and station capacity. This observation validates our model's focus on non-monetary costs as the dominant drivers of route choice for the average commuter in this system, ensuring the model's parameters accurately reflect the key trade-offs they prioritize.

Table 10. Monthly observed and model calculated route choice probabilities and total sample sizes for Musashi-Shinjō to Shinjuku OD pair.

Month	R1	R2	R3	R4	Sample Size
1	0.56	0.21	0.10	0.13	30
2	0.57	0.20	0.12	0.11	60
3	0.34	0.28	0.12	0.26	57
4	0.53	0.18	0.13	0.16	60
5	0.42	0.15	0.13	0.30	58
6	0.39	0.22	0.14	0.25	57
7	0.50	0.16	0.13	0.21	73
8	0.31	0.27	0.16	0.26	62
9	0.44	0.18	0.15	0.23	65
10	0.32	0.30	0.24	0.14	64
11	0.35	0.22	0.16	0.27	60
12	0.56	0.17	0.10	0.17	55
Model Predicted (Model 1 variants)	0.32	0.27	0.18	0.23	

Note: To minimize the influence of individual commuting strategies, this origin-destination (OD) pair was selected because commuters almost exclusively travel during peak hours, ensuring the peak dummy variable (PK) is consistently equal to 1.

Additional cases are shown in the Supplement Figures S2 and S3, including scenarios where people choose less crowded transfer stations despite slightly longer travel times, and cases where people tolerate extra transfers for less crowded lines.

Consequently, the model's predicted route usage is more evenly distributed across all options compared to the highly concentrated usage observed in raw GPS data, where minor behavioral factors not captured by the model lead to more dispersed choices.

Despite this moderate uncertainty, the model successfully captures the fundamental ranking of alternatives, as evidenced by the high Spearman correlation. This indicates that while we may not perfectly predict the exact probability for each route, the model reliably identifies which routes are more or less likely to be chosen: the essential information for transportation planning and network optimization.

5.2. OD-Specific Parameter Estimation: Systematic Heterogeneity

Following the aggregate analysis, we apply the same model structure to individual OD pairs to examine heterogeneity and address: (1) whether OD pairs can be grouped by distinct choice patterns (Section 5.2); and (2) which OD-pair characteristics drive these differences, including the time-equivalent cost of transfers (Sections 5.2.1 and 5.2.2).

Table 11 summarizes the Model 2(OD-specific estimation) estimation results alongside the Model 1(city-wide pooled estimation) aggregate estimates from Table 7, providing a unified comparison of both model variants. The probability density functions of the OD-specific parameters (Figure 9b–g) reveal that preference variation across OD pairs is systematic rather than random. Most parameters exhibit unimodal distributions whose primary modes align closely with the Model 1 aggregate estimates (indicated by blue dotted lines in Figure 9), confirming that the city-wide parameters represent the central tendency of OD-level behavior. Because both Model 1 and Model 2 estimates share the same $[0, 1]$ normalization (Section 4.3.1), β^{OD} values are directly comparable to the city-wide benchmark $\beta_{city} = 1.658$: $\beta^{OD} > \beta_{city}$ indicates more deterministic choices than the metropolitan average,

$\beta^{OD} < \beta_{\text{city}}$ indicates greater stochasticity (Figure 9). The same convention enables standardized cross-city comparison once comparable datasets become available.

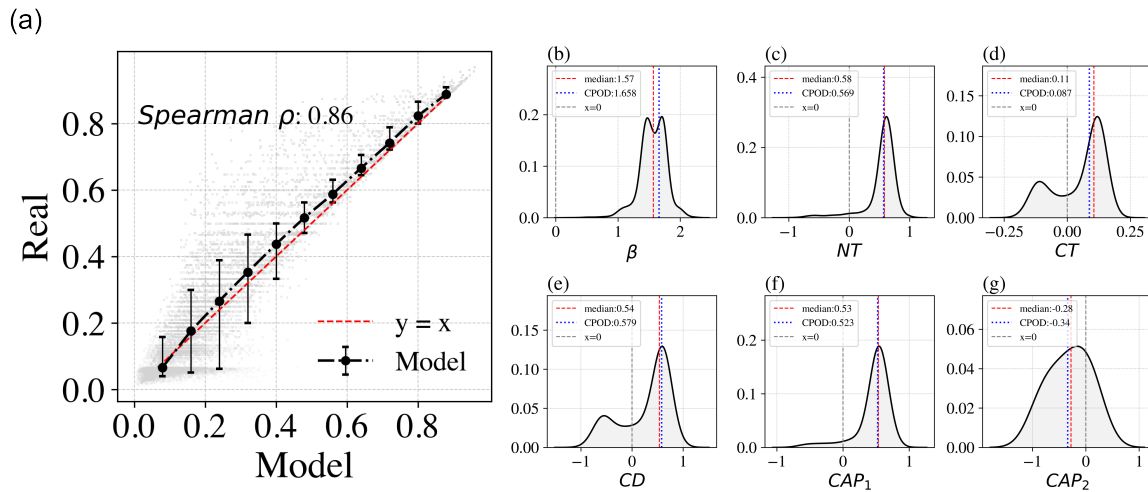


Figure 9. OD-specific model performance and parameter heterogeneity. **(a)** Model validation for OD-specific estimations showing predicted versus observed route choice probabilities. Error bars represent 25th-75th percentiles; the red dashed line indicates perfect prediction ($y = x$); Spearman $\rho = 0.86$. **(b–g)** Probability density functions of estimated parameters across all OD pairs: **(b)** inverse of the unobserved error scale parameter β ; **(c)** number of transfers (NT) coefficient; **(d)** commuting time coefficient (CT); **(e)** commuting distance coefficient (CD); **(f)** first transfer capacity coefficient (CAP_1); and **(g)** second transfer capacity coefficient (CAP_2). Blue dotted lines indicate the aggregate Model 1 (city-wide pooled estimation) estimates from Table 7, confirming that the city-wide parameters align with the primary mode of the OD-specific distributions, red vertical lines indicate median values; grey vertical lines indicate zero. The distributions reveal systematic heterogeneity in route choice preferences across different origin-destination pairs.

Table 11. Comparison of Model 1 and Model 2 estimation results.

Parameter	Model 1 Estimate	Model 2	
		Mean	Median
Scale parameter (β)	1.658	1.561	1.567
No. of Transfers (NT_{nj})	0.569 ***	0.536	0.580
Commuting Time (CT_{nj} , min)	0.087 ***	0.048	0.105
Commuting Distance ($\log(CD_{nj})$, km)	0.579 ***	0.262	0.535
Peak Departure (PK_{nj} , dummy)	−0.117 ***	—	—
Capacity at 1st transfer ($\log(CAP_{nj1})$)	0.523 *	0.465	0.533
Capacity at 2nd transfer ($\log(CAP_{nj2})$)	−0.340 ***	−0.312	−0.277
		Model 1	Model 2 (median)
Number of OD pairs		5,908	5,908
McFadden's R^2		0.232	0.383
Hit rate		55.8%	71.0%
Spearman ρ		0.724	0.860
% of ODs with $R^2 > 0.232$		—	74.8%
% of ODs with hit rate $> 55.8\%$		—	73.8%

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. The Peak Departure variable (PK) is not estimated in the Model 2 model because the Model 2 analysis separates peak and off-peak commuters into distinct sub-samples.

However, the distributions for the inverse of the unobserved error scale parameter (β) and the coefficients for Commuting Time (CT) and Commuting Distance (CD) display distinct bimodal patterns (Figure 9b,d,e). This bimodality is a key finding: it reveals the existence of structurally distinct behavioral groups within the network, rather than a continuum of idiosyncratic variation.

Specifically, while the main peak for CT and CD coefficients is positive (aligning with the aggregate result, where greater time or distance increases perceived cost), a substantial subset of OD pairs exhibits negative coeffi-

cients. In the context of our energy-cost model, a negative cost coefficient for time or distance indicates that, for these specific corridors, longer times or distances are associated with lower perceived energy cost. This counter-intuitive result immediately establishes that two structurally distinct OD groups coexist in how commuting time and distance enter the realized choice distribution; their behavioral interpretation will be examined in Section 5.2.2. The heterogeneity in β , meanwhile, will be explored in Section 5.2.1 to account for temporal differences between peak and off-peak departures. Here, systematic heterogeneity refers strictly to the variation in commuter preferences for a single route attribute across different origin-destination pairs.

To ensure that these bimodal patterns are not artifacts of limited observations in certain OD pairs, we verified the statistical adequacy of our observation-to-parameter ratio and conducted a robustness analysis restricting the sample to OD pairs with at least 40 commuters. As detailed in the Supplementary Material (Robustness Check, Figure S8), the bimodal structures remained consistent, confirming the robustness of the finding.

The Model 2 (OD-specific estimation) model's predictive performance confirms the reliability of these OD-level estimates: as shown in Table 11, the median McFadden's R^2 is 0.383 and the median hit rate is 71.0%, with 74.8% and 73.8% of OD pairs exceeding the corresponding Model 1(city-wide pooled estimation) benchmarks, respectively (see Supplementary Material Figure S12 for distributions).

5.2.1. Temporal Dimension: Peak versus Off-Peak Dynamics

Analysis of the OD-specific parameters separated by departure time (peak versus off-peak) reveals distinct behavioral dynamics, which concurrently explains the bimodal distribution observed for the inverse of the unobserved error scale parameter (β) in Figure 9b.

Figure 10a shows a clear distinction in the inverse of the unobserved error scale parameter (β). Peak-hour commuters exhibit lower β values (median $\simeq 1.52$), indicating less deterministic and more exploratory choice behavior. This reduction in determinism can be attributed to two main factors: (1) strategic Exploration: Commuters actively deviate from minimal-cost routes to navigate severe crowding and capacity constraints, often prioritizing securing a seat or avoiding specific congested points; (2) peak-hour operation is more susceptible to service disruptions and delays, which introduce higher variance in actual commuting times. As shown in Figure 10c, peak-hour commuting time exhibits more negative parameters compared to off-peak hours, indicating that for more OD pairs, longer routes are taken more frequently during peak periods. This counterintuitive pattern provides evidence that unexpected factors (such as crowding and delays) make peak-hour commuting less predictable. The underlying reasons for these negative time parameters are discussed further in the following section.

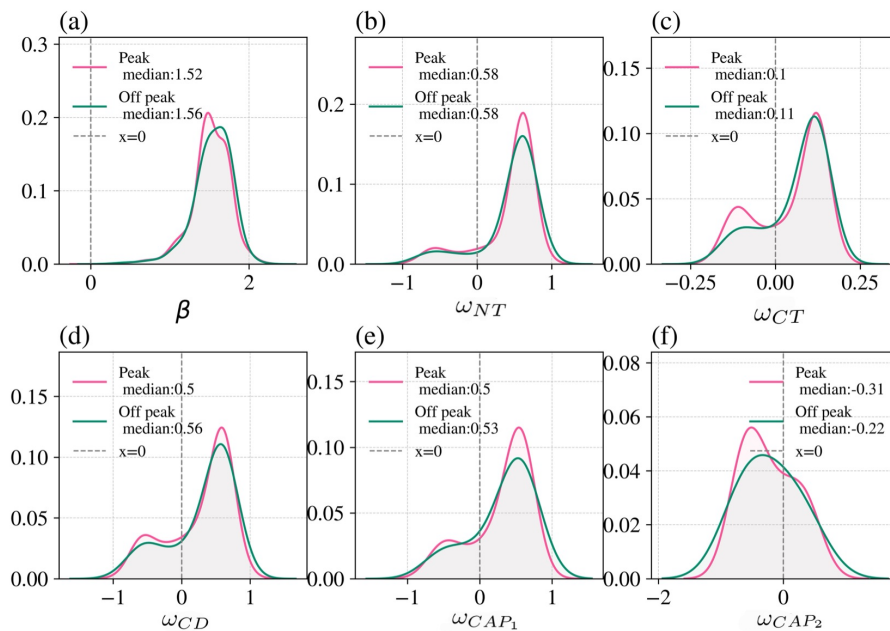


Figure 10. Comparison of parameter distributions between peak and off-peak periods. Probability density functions of estimated parameters across OD pairs, separated by departure time: pink lines show peak-hour distributions; green lines show off-peak distributions. (a) Scale parameter β ; (b) number of transfers coefficient (NT); (c) commuting time coefficient (CT); (d) commuting distance coefficient (CD); (e) first transfer capacity coefficient (CAP_1); (f) second transfer capacity coefficient (CAP_2). Grey dashed lines indicate zero reference.

Off-Peak Commuters: Exhibit higher β values (median ≈ 1.56), reflecting more deterministic behavior. With reduced crowding and greater capacity, the need for strategic deviation is minimized. Furthermore, the train operation schedule is generally more stable during off-peak hours, leading to lower time variance and allowing commuters to more consistently select routes based on minimized cost or habit.

The remaining route attributes (Figure 10b–f) exhibit similar central tendencies between peak and off-peak periods. This finding suggests that fundamental cost, such as the inherent aversion to transfers, time, and distance, remain consistent regardless of departure time. Consequently, for the subsequent detailed analysis of the bimodal heterogeneity observed in the Commuting Time and Commuting Distance coefficients (Figure 9c,d), we proceed without separating peak and off-peak commuters, focusing instead on other contextual factors that drive these systematic differences.

5.2.2. Explaining Counterintuitive Parameters: OD Pair Characteristics

The systematic heterogeneity revealed in the OD-specific model, particularly the bimodal distributions of the Commuting Time (CT) and Commuting Distance (CD) coefficients, presents a crucial puzzle: why do some OD pairs exhibit negative coefficients, suggesting commuters prefer routes with longer times or distances?

To investigate this phenomenon, we segregated OD pairs based on the sign of their CT and CD coefficients and analyzed three key characteristics: (1) the central tendency of route attribute values (time/distance); (2) the distributional shape through skewness of within-OD attribute distributions; and (3) the frequency of transfers (Table 4). Figure 11 presents comparative analyses ((a)–(f)) and representative examples through violin plots ((g)–(h)). A sensitivity analysis (Supplementary Section Sensitivity Analysis) confirms that this heterogeneity persists even when restricting the dataset to OD pairs with 40 or more commuters, ruling out small-sample variance as the primary cause.

Commuting Time Heterogeneity

Why do commuters appear to “prefer” longer routes? Railway network service constraints and crowd-avoidance behavior drive this pattern. The negative commuting time (CT) coefficients observed in some OD pairs do not imply a genuine behavioral preference for longer travel times. Instead, this apparent anomaly is a statistical artifact driven by the underlying travel-time distributions. When unobserved utility trade-offs (e.g., choosing slower local trains to secure seating) or structural network constraints (e.g., fewer headways, longer transfer waiting times) cause a dominant subset of commuters to cluster at higher time values, the MNL regression mechanically yields a negative CT coefficient.

We establish this mechanical reversal through three interconnected observations. First, negative-coefficient ODs systematically involve more complex route profiles with longer journeys (median 42.0 vs. 31.0 min) and a higher reliance on transfers, compared to positive-coefficient ODs (Figure 11a,c), indicating greater reliance on interruption-heavy network structures. Second, realized commuting times in these ODs are frequently bimodal. Rather than concentrating near the minimum possible time, the probability mass shifts toward the upper travel-time bound, reflecting distinct route choice clusters within the same OD pair (Figure 11b,g,h). Finally, this structural bimodality induces a mechanical sign reversal. Individual commuters are still maximizing their overall utility by incorporating unobserved comforts or avoiding transfer uncertainties. However, because their chosen commutes cluster at the longer-time mode of the bimodal distribution, the standard MNL regression mechanically assigns a negative parameter to commuting time.

Commuting Distance Heterogeneity

Distance coefficients exhibit parallel patterns. Negative-coefficient ODs correspond to longer journeys (median: 16.9 km vs. 11.8 km, Figure 11d) with lower skewness (Figure 11e), mirroring the time analysis. Transfer complexity follows the same pattern (Figure 11f): fewer direct travelers, more multi-transfer users.

Importantly, commuters may occasionally prefer geometrically longer routes if those specific alternatives happen to offer express services, higher service frequency, or more reliable connections compared to the more direct paths. Network topology constrains certain OD pairs into choice sets where preferred (lower perceived-cost) alternatives are not necessarily the shortest.

Mechanisms Underlying Bimodal Distributions

Bimodal patterns in both time and distance arise from two structural mechanisms, complemented by a behavioral explanation. First, transfers function as trip interruptions analogous to traffic signals. Previous research [59] showed traffic signals induce bimodal travel time distributions; our analysis reveals this extends to railway commut-

ing. A possible explanation is that service frequency shapes these distributions: with long headways, commuters face binary outcomes—those arriving just before departure board immediately, while those just missing a train wait prolonged periods. This waiting-time clustering introduces significant delays and variance (Figure 11c,f), explaining why high-transfer ODs develop bimodal characteristics concentrated at longer times and distances.

Second, even without transfers ($n = 0$), bimodality emerges from the coexistence of local and express train services. For longer-distance OD pairs, the time and distance differentials between local trains (all stops) and express trains (limited stops) increase proportionally with journey length, creating two distinct travel-time clusters that intensify with distance. Beyond these structural mechanisms, travel comfort offers a complementary behavioral explanation: commuters on long-distance routes may prefer slower direct services that offer greater seating availability over faster but less predictable transfer-dependent alternatives, further reinforcing the separation between the two modes.

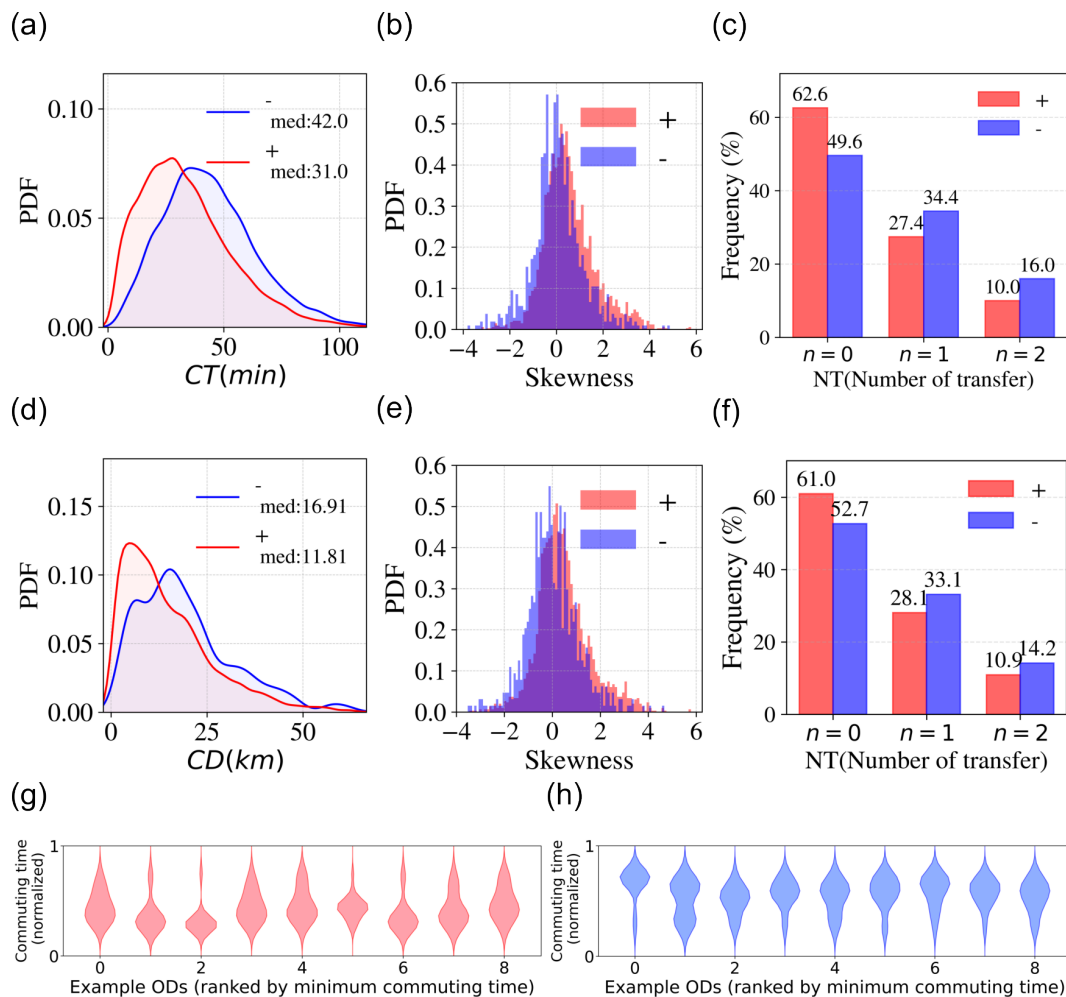


Figure 11. Characterization of OD pairs with positive versus negative commuting time and distance coefficients. (a) Probability density distribution of commuting time (CT) for OD pairs with positive (red, “+”) versus negative (blue, “-”) coefficients, with median values shown in legends. (b) Distribution of skewness values for within-OD commuting time distributions, separated by coefficient sign (red “+” and blue “-”). (c) Frequency distribution of number of transfers (NT) for CT-positive versus CT-negative OD pairs, with percentages indicating the proportion of ODs in each category. (d) Probability density distribution of commuting distance (CD) for OD pairs with positive versus negative coefficients, with median values shown in legends. (e) Distribution of skewness values for within-OD commuting distance distributions. (f) Frequency distribution of number of transfers for CD-positive versus CD-negative OD pairs. (g,h) Violin plots showing commuting time distributions for route alternatives within nine representative OD pairs, with commuting time rescaled to the unit interval [0,1] via min-max normalization (0 represents the minimum commuting time, 1 represents the maximum commuting time within each OD pair), ordered from left to right by minimum commuting time: (g) OD pairs with positive CT coefficients exhibit concentrated distributions near lower normalized values, while (h) OD pairs with negative CT coefficients show distributions with peaks shifted toward longer travel times.

Quantifying Transfer Penalty

By analyzing the ratio of model coefficients, we can quantify the implicit trade-offs in commuters' route selection. Specifically, the transfer penalty expressed in time-equivalent units can be calculated as:

$$\text{Transfer Penalty (minutes)} = \frac{\omega_{NT}}{\omega_{CT}}, \tag{14}$$

where ω_{NT} is the coefficient for number of transfers and ω_{CT} is the coefficient for commuting time. This ratio quantifies the energy cost of transfers in travel time equivalents. Dimensionally, dividing the transfer coefficient (energy cost per transfer) by the time coefficient (energy cost per minute) cancels the implicit energy unit, translating the transfer burden directly into an equivalent value of commuting minutes.

Using the city-wide coefficients in Section 5.1, Table 7 ($\omega_{NT} = 0.569$, $\omega_{CT} = 0.087$), Equation ((14)) yields an average transfer penalty of approximately 7 min. In other words, a typical Tokyo commuter treats one transfer as imposing a mental and physical cost equivalent to 7 extra minutes of in-vehicle time, and will only switch from a direct route to a one-transfer alternative when the latter saves at least 7 min.

To examine spatial variations, we apply Equation (14) at the OD level. As detailed in Table 12, this calculation is strictly performed on the final subset of 2,557 OD pairs that admit transfer alternatives and exhibit statistically significant positive coefficients for both time and transfers. Aggregating these OD-specific penalties by commuting distance (Figure 12) reveals that the 7-min city-wide average conceals substantial distance dependence. For short-to medium-distance commutes (<50 km), penalties cluster below 10 min, close to the Model 1(city-wide pooled estimation) baseline. Between 50 and 70 km, the penalty steadily rises to ~10–15 min. While the 80 km bin exhibits a sharp peak near 30 min, its wide confidence intervals suggest high volatility. Beyond this point, the penalty declines for the longest trips, where data becomes sparse. This escalation toward the periphery likely reflects the heightened inconvenience of transfers for suburban commuters, who often face longer headways and less seat availability at peripheral stations; this increase aligns with the spatial heterogeneity observed in [60].

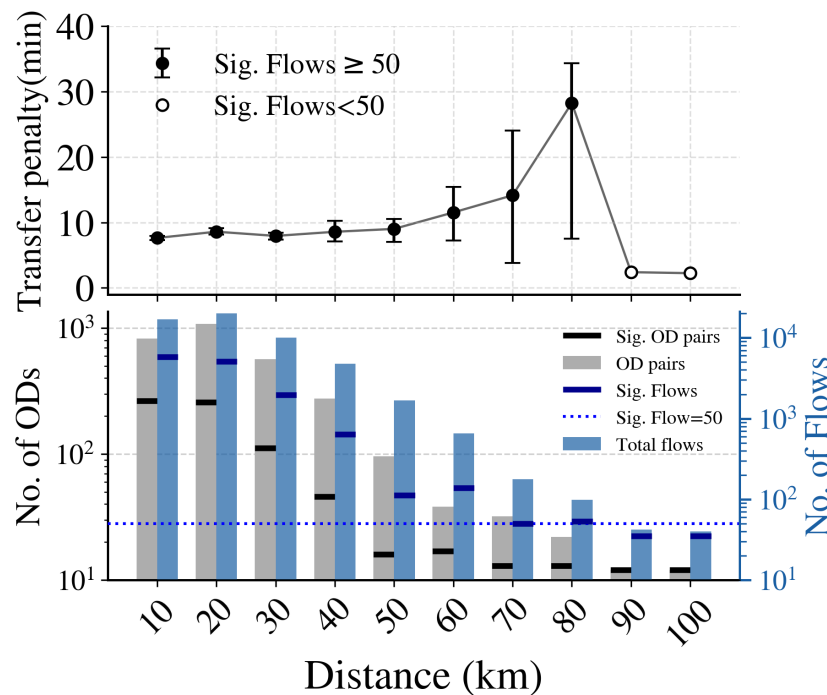


Figure 12. Transfer penalty as a function of commuting distance. Based on the 2557 jointly significant OD pairs identified in Table 12. Top: Per-OD transfer penalty ω_{NT}/ω_{CT} summarized by the bin-wise median; error bars are 95% percentile bootstrap confidence intervals of the median (10,000 resamples). Filled markers indicate bins with at least 50 significant flows; open markers indicate bins below this threshold, which should be interpreted with caution. Bottom: Number of OD pairs (grey, left axis) and total trajectories (blue, right axis) per distance bin. The blue dotted line marks the 50-flow threshold separating filled and open markers in the top panel.

Our estimated transfer penalty (approximately 7 min) is substantially lower than the 15–18 min reported by Garcia-Martinez et al. (2018) [18] for Madrid. There are two potential reasons for this discrepancy.

First, in our state-determination procedure, each 1 km grid cell is treated as a single state. Consequently, the

intra-cell walking time, specifically the duration from entering the grid to reaching the exact physical destination, is inherently omitted. Estimating the travel time coefficient ω_{CT} using this undercounted travel time introduces an upward bias in ω_{CT} , thereby compressing the penalty ratio ω_{NT}/ω_{CT} .

Second, stated-preference (SP) and revealed-preference (RP) estimates capture distinct constructs. The survey-based 15–18 min penalty reflects subjective psychological resistance (e.g., *anxiety*). However, as Garcia-Martinez et al. [18] themselves note, habitual transferring confers a positive utility that offsets this resistance. Our GPS-based revealed-preference estimates capture habituated commuters who have already internalized these psychological costs, yielding a lower, behaviorally realized penalty.

Table 12. Identification quality of OD-level estimates and parameter regimes

Criterion	<i>N</i>	% of Total OD Pairs 5908
All ODs with Model 2 estimates	5908	100.0
Without transfer alternatives	1457	24.7
With transfer alternatives	4451	75.3
Criterion	<i>N</i>	% of 4451
For ODs with transfer alternatives		
Positive regime ($\omega_{NT} > 0$ and $\omega_{CT} > 0$)	3653	82.1
Negative regime (at least one ≤ 0)	798	17.9
Criterion	<i>N</i>	% of 3653
Statistical significance (positive regime only for calculating transfer penalty):		
ω_{CT} significantly > 0 ($p < 0.05$, one-sided)	2559	70.0
ω_{NT} significantly > 0 ($p < 0.05$, one-sided)	3364	92.1
Both significantly > 0	2557	70.0

Note: This table illustrates the sequential filtering of OD pairs. Percentages in each section are calculated relative to the subset defined in the preceding section, as indicated by the column headers. The “positive regime” is defined by point estimates where both transfer (ω_{NT}) and time (ω_{CT}) coefficients are strictly positive. Statistical significance is evaluated via the asymptotic Wald test applied separately to each coefficient, with one-sided p -values computed under $H_0 : \omega \leq 0$ against $H_1 : \omega > 0$.

6. Conclusions and Limitations

This study addresses the central research question: Do commuters exhibit consistent route choice “rationality” across different contexts and time in Tokyo’s morning commute? The analysis reveals that Tokyo railway commuters exhibit strong but non-deterministic preference, with route choices clustering into highly preferred or highly unlikely alternatives with few intermediate options. These choices are driven primarily by factors including travel time, distance, and transfers. Additionally, commuters tend to avoid high-capacity stations for their first (or only) transfer, but choose high-capacity stations for their second transfer. Importantly, the results demonstrate nuanced consistency: the aggregate level of choice determinism and route attribute preferences remain temporally stable across seasons, showing structural consistency in decision-making.

Furthermore, the OD-pair-based heterogeneity analysis reveals significant contextual variation. Peak-hour commuters are less deterministic (indicated by lower β), exhibiting more exploratory behavior as they strategically deviate from minimal-cost routes to secure seats and ensure punctual arrival under capacity constraints. In contrast, off-peak commuters show more deterministic choices. Critically, we identify transfer interruptions as a key factor introducing behavioral heterogeneity and choice uncertainty. The disaggregate analysis confirms that this variation is driven primarily by network structure: OD pairs with high transfer complexity display counterintuitive preferences (choosing longer routes). Transfers create bimodal travel time distributions, where some commuters arrive significantly earlier while others arrive later [59], generating psychological pressure of *I could have arrived much earlier if I had caught the train leaving 1 minute before*. This structural uncertainty, amplified by the high transfer penalty (7-min equivalent), fundamentally limits model predictability for transfer-intensive routes [60]. These findings emphasize that policymakers must exercise greater caution in transfer route design, as transfer complexity drives both commuter stress and behavioral unpredictability in dense urban railway networks.

Methodologically, we demonstrate the successful application of this framework to a large-scale, passively collected GPS dataset. We addressed the technical challenges of processing raw smartphone GPS data, including its variable spatial resolution and noisy trajectories, by developing a comprehensive processing pipeline. This pipeline

successfully transforms raw location records into high-resolution commuting mode classifications (distinguishing between railway, walking, cycling, and vehicle-based travel) [28,38] and specific route trajectories. Crucially, we overcame the computational challenge of route identification for large-scale GPS data [7] by developing a novel spatial analysis method that accurately reconstructs chosen railway paths through transfer station identification.

Theoretically, our study provides strong empirical validation that the complex route choice system of Tokyo's railway network operates in a generally stable state. By confirming temporal consistency across periods, we demonstrate that the current transportation system maintains stable collective behavior even as individual users make personal trade-offs [3,4,61].

Notably, the finding that peak-hour commuters exhibit lower determinism is consistent with previous studies suggesting that these commuters adopt more strategic behaviors when choosing routes [5], deviating from pure minimal-cost paths to secure seats or ensure punctual arrival under strict capacity constraints [19]. We contribute to this literature by providing a quantitative metric to measure this strategic shift.

Most importantly, the observed heterogeneity offers profound implications for future modeling studies. While the majority of route choice behaviors are predictable, we find that routes characterized by long travel times, long distances, and frequent transfers are a primary source of modeling difficulty, often being misclassified as unexplainable heterogeneity. Our results complement previous studies on vehicle travel which found that trip interruptions [59], specifically traffic signals, induce bimodal distributions in commuting time when origin and destination locations are fixed. Our analysis reveals that this structural pattern is mirrored in long time and distance railway commuting: transfers act analogously to traffic signals as trip interruptions, inducing similar bimodal distributions in route choice behavior. This suggests that future route choice models should account for the bimodal nature of interruption-related uncertainty rather than assuming standard symmetric distributions.

6.1. Connection to Maximum Entropy Inference

The MNL model admits a well-established mathematical correspondence with the canonical ensemble of statistical mechanics: available routes correspond to microstates, generalized cost to energy, and the logit choice probability emerges as the maximum-entropy distribution subject to an observed expected-cost constraint [30,52,62]. As detailed in Appendix B and Supplementary Section 16, the equivalence is algebraic: the MaxEnt approach maximizes Shannon entropy subject to the constraint that model-predicted expected route costs match their observed aggregate values, while the MLE approach maximizes the log-likelihood of observed choices under the MNL functional form.

Despite their different starting points—one information-theoretic, the other statistical—both yield the same first-order condition and therefore the same parameter estimates [52,62]. What differs is interpretation: in the econometric reading, ω_{raw} is the coefficient vector that best explains individual choices; in the MaxEnt reading, it is the Lagrange multiplier that enforces the macroscopic cost constraint. The logsum term likewise admits dual meaning—the log-partition function in statistical mechanics, and the expected maximum utility in discrete choice theory [15]. Helbing & Nagel [30] provide a representative review of this duality in the physics-of-cities literature. Building on this theoretical duality, the present manuscript establishes the empirical and methodological groundwork—a validated, large-scale route choice dataset with stable aggregate parameters, that enables future Maximum Entropy-based investigations.

6.2. Limitations

Despite its significant contributions, this study has several limitations. Fare data were not collected, which prevents direct estimation of monetary trade-offs in route choice. Furthermore, the multi-stage methodology, from identifying transportation modes and transfer stations to reconstructing final chosen routes from raw GPS data, introduces potential accuracy issues at each step, albeit minimal. A related limitation is the lack of semantic labels, meaning some brief interruptions could represent non-transit activities like buying breakfast. However, strict temporal and spatial filtering in our methodology minimizes this noise. Combined with the strong cultural norm of direct workplace travel during Tokyo's morning peak, these rare misclassifications are unlikely to skew the aggregate route choice models.

While travel time components such as access, egress, walking, and waiting times are known to influence commuter valuation, they were not explicitly calculated in this study. As illustrated in Supplementary Figure S7, the GPS logging frequency provided by Agoop is insufficiently dense to pinpoint precise state changes at this granular level. Specifically, there is typically a spatial and temporal gap between the final point of an *access* walk and the first logged *move* point on a train. According to the data provider's handbook, the dataset most reliably captures trip initiations, movement state changes, or transitions when users emerge from underground for transfers.

Consequently, our model focuses on the most robustly captured attributes: total door-to-door commuting time, high-speed movement points for network matching, and *stroll* points to identify transfer behavior.

Furthermore, our reliance on door-to-door time means the model cannot explicitly differentiate between access and egress modes (e.g., walking, cycling, or driving). While these modes influence total travel time, treating these segments uniformly is a robust approximation for the Tokyo metropolitan context. According to the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) [63], 87% of regional commuters access stations via active transport (70% walking, 17% bicycles). Motorized access accounts for only 13% due to parking constraints and high bus fares. However, because our model calculates a route's base time using the median duration of all commuters observed on that path, grouping these access modes introduces measurement bias. Specifically, a high proportion of cyclists on a given route will skew the median travel time lower, making distant stations appear more accessible than they actually are for pedestrians. Distinguishing access modes to calculate mode-specific travel times represents an important direction for future refinement.

Additionally, residual correlations may remain among route categories that share physical infrastructure. Future research should explicitly account for this overlap. Following Dixit et al. [51], who demonstrated that shared transfer nodes drive traveler perception in transit networks, integrating a Path Size Correction Logit (PSCL) represents a promising methodological extension.

Furthermore, to mitigate the inherent multicollinearity between travel time and distance, future model specifications should replace raw distance with distance circuitry. As demonstrated by Dixit et al. [51], circuitry effectively captures the penalty of indirect routing independent of travel time.

Additionally, train frequency was omitted due to unavailable unified timetables, though Tokyo's short headways minimize this impact. Finally, station capacity is treated as a static annual maximum in the current model. A more refined approach would assign time-specific capacity values matched to each commuter's actual transit time at each transfer station. We leave this as a direction for future work.

Nonetheless, the study makes important contributions: this is the first attempt to apply a mobility dataset of this unprecedented scale and spatial resolution to railway route choice modeling, yielding significant empirical findings. Crucially, the framework provides an interpretable, theory-driven approach, allowing us to directly interpret and quantify commuter behavior based on observable route attributes.

Supplementary Materials

The following supporting information can be downloaded at: <https://media.sciltp.com/articles/others/2606231539173202/J.-Soc.-Phys.-25100061-Author-Supplementary.pdf>. Reference [64] is cited in supplementary materials.

Author Contributions

Y.Y.Z.: designed the research plan, developed methods of data analysis, performed the numerical calculations, and wrote the manuscript. H.T.: checked methods of data analysis and revised the manuscript. M.T.: led this project, and directed writing of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (B) (GrantNumber 23K22980 to MT). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Institutional Review Board Statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements.

Informed Consent Statement

Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Data Availability Statement

The GPS data cannot be shared publicly because data is available only on request from a third party. Data are available from Agoop Corporation, a Japanese private company that provides location information big data acquired from smartphone applications. The specific product is "Pointo-gata ryoudou-jinkou data"

(Point-type population data). Interested researchers who meet the criteria for access to confidential data can visit <https://agoop.co.jp/service/dynamic-population-data/> for more information. Spatial data of the physical railway network were obtained from the National Land Numerical Information (N02, 2023) dataset provided by Japan's Ministry of Land, Infrastructure, Transport and Tourism (MLIT). The data are freely available under the CC BY 4.0 license at <https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N02-2023.html>.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used AI-assisted writing tools to check for grammatical errors and improve the English language style. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix A. Mathematical Derivations for Parameter Normalization

This appendix provides detailed mathematical derivations supporting the Delta method calculations described in Section 4.3 of the main text.

- (1) Estimate Raw Parameters: Use Maximum Likelihood Estimation (MLE) to find the unconstrained weight vector, $\hat{\omega}_{\text{raw}}$.
- (2) Calculate Raw Covariance: Approximate the covariance matrix of these raw parameters, $\hat{\Sigma}_{\omega_{\text{raw}}}$, by inverting the Hessian matrix of the negative log-likelihood function.
- (3) Apply Delta Method: In order to get the variance and covariance of the non-linearly transformed parameters (β, ω) , use the Delta method to calculate them so that it can directly get from the ω_{raw} , whose covariance $\hat{\Sigma}_{\omega_{\text{raw}}}$ can be approximated MLE during optimization process. This gives:
 - The variance of the scale parameter, $\text{Var}(\hat{\beta})$.
 - The full covariance matrix of the normalized weights, $\text{Cov}(\hat{\omega})$.
- (4) Test for Consistency: Using these variances and covariances, we perform statistical tests (like the Wald test) and construct confidence intervals to determine if the solutions for β and ω are consistent across the 12 months.

To maintain consistency, we use the same notation as the main text: ω_{raw} for the unconstrained weight vector, β for the scale parameter, and ω for the normalized weight vector.

Appendix A.1. Gradient and Jacobian Derivations

Appendix A.1.1. Gradient of the Norm $\beta = \|\omega_{\text{raw}}\|_2$

Let $f(\omega_{\text{raw}}) = \|\omega_{\text{raw}}\|_2 = \left(\sum_{j=1}^k \omega_{j,\text{raw}}^2\right)^{1/2}$. The gradient is a vector of partial derivatives with respect to each element of ω_{raw} :

$$\nabla_{\omega_{\text{raw}}} f = \begin{bmatrix} \frac{\partial f}{\partial \omega_{1,\text{raw}}} \\ \frac{\partial f}{\partial \omega_{2,\text{raw}}} \\ \vdots \\ \frac{\partial f}{\partial \omega_{k,\text{raw}}} \end{bmatrix} \quad (\text{A1})$$

We compute the partial derivative for a single element $\omega_{i,\text{raw}}$ using the chain rule:

$$\begin{aligned} \frac{\partial f}{\partial \omega_{i,\text{raw}}} &= \frac{\partial}{\partial \omega_{i,\text{raw}}} \left(\sum_{j=1}^k \omega_{j,\text{raw}}^2 \right)^{1/2} \\ &= \frac{1}{2} \left(\sum_{j=1}^k \omega_{j,\text{raw}}^2 \right)^{-1/2} \cdot \frac{\partial}{\partial \omega_{i,\text{raw}}} \left(\sum_{j=1}^k \omega_{j,\text{raw}}^2 \right) \\ &= \frac{1}{2f} \cdot (2\omega_{i,\text{raw}}) = \frac{\omega_{i,\text{raw}}}{f} \end{aligned}$$

Since this result holds for every element, the full gradient vector is:

$$\nabla_{\omega_{\text{raw}}} f = \frac{1}{f} \begin{bmatrix} \omega_{1,\text{raw}} \\ \omega_{2,\text{raw}} \\ \vdots \\ \omega_{k,\text{raw}} \end{bmatrix} = \frac{\omega_{\text{raw}}}{f} = \frac{\omega_{\text{raw}}}{\|\omega_{\text{raw}}\|_2} = \omega \quad (\text{A2})$$

This demonstrates that the gradient of the norm with respect to the raw weight vector equals the normalized weight vector itself.

Appendix A.1.2. Jacobian of the Normalization $\omega = \omega_{\text{raw}} / \|\omega_{\text{raw}}\|_2$

Applying the quotient rule for vector functions:

$$d\omega = d\left(\frac{\omega_{\text{raw}}}{f}\right) = \frac{1}{f}d\omega_{\text{raw}} + \omega_{\text{raw}} \cdot d\left(\frac{1}{f}\right) \quad (\text{A3})$$

The second term is derived from the chain rule:

$$d\left(\frac{1}{f}\right) = -\frac{1}{f^2}df = -\frac{\omega_{\text{raw}}^\top d\omega_{\text{raw}}}{f^3} \quad (\text{A4})$$

Substituting this back and simplifying, we get:

$$\begin{aligned} d\omega &= \frac{1}{f}d\omega_{\text{raw}} - \frac{\omega_{\text{raw}}(\omega_{\text{raw}}^\top d\omega_{\text{raw}})}{f^3} \\ &= \frac{1}{f} \left(I - \frac{\omega_{\text{raw}}\omega_{\text{raw}}^\top}{f^2} \right) d\omega_{\text{raw}} \\ &= \frac{1}{\beta} (I - \omega\omega^\top) d\omega_{\text{raw}} \end{aligned} \quad (\text{A5})$$

Therefore, the Jacobian matrix is:

$$\mathbf{J}_\omega = \frac{\partial \omega}{\partial \omega_{\text{raw}}} = \frac{1}{\beta} (I - \omega\omega^\top) \quad (\text{A6})$$

The matrix $(I - \omega\omega^\top)$ is a projection matrix that maps any vector onto the space orthogonal to ω , which is crucial for maintaining the unit length of the normalized vector under small perturbations.

Appendix A.2. Delta Method for Standard Errors

The Delta method is a powerful statistical technique used to approximate the variance and covariance of a function of an asymptotically normal random variable. In simpler terms, if the variance of a parameter estimate is known, the Delta method helps to find the variance of a new parameter that is a nonlinear function of the original one.

The core of the Delta method is the first-order Taylor expansion, which approximates a nonlinear function with a linear one. We start with a vector of random variables, \mathbf{X} , with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We want to find the approximate covariance of a new vector, $g(\mathbf{X})$, which is a nonlinear function of \mathbf{X} .

The Taylor expansion of $g(\mathbf{X})$ around its mean $\boldsymbol{\mu}$ is:

$$g(\mathbf{X}) \approx g(\boldsymbol{\mu}) + \mathbf{J}(\mathbf{X} - \boldsymbol{\mu}) \quad (\text{A7})$$

where \mathbf{J} is the Jacobian matrix of g evaluated at the mean $\boldsymbol{\mu}$. It is a matrix of all the first-order partial derivatives of the function g . This Jacobian matrix is the linear part of the approximation, capturing how a small change in \mathbf{X} affects the output of $g(\mathbf{X})$.

Appendix A.2.1. General Framework

Given that the raw MLE parameter estimates $\hat{\omega}_{\text{raw}}$ are asymptotically normal with covariance matrix $\hat{\boldsymbol{\Sigma}}_{\omega_{\text{raw}}}$, the Delta method provides the covariance matrix of a transformed vector $g(\hat{\omega}_{\text{raw}})$ as:

$$\text{Cov}(g(\hat{\omega}_{\text{raw}})) \approx \mathbf{J} \hat{\boldsymbol{\Sigma}}_{\omega_{\text{raw}}} \mathbf{J}^\top \quad (\text{A8})$$

where \mathbf{J} is the Jacobian of the transformation g evaluated at $\hat{\omega}_{\text{raw}}$.

Here, $g(\omega_{\text{raw}})$ represents the nonlinear transformation that normalizes ω_{raw} and extracts the scale parameter β .

- $g(\omega_{\text{raw}})$ is the new vector $\theta_{\text{final}} = (\beta, \omega)^T$.
- This transformation is nonlinear because it involves a square root (in the norm calculation) and division by a variable quantity (in the normalization).

Variance of the Scale Parameter β

The scale parameter is $\beta = g_1(\omega_{\text{raw}}) = \|\omega_{\text{raw}}\|_2$. Its gradient is the first row of the full Jacobian matrix.

$$\text{Var}(\hat{\beta}) \approx (\nabla_{\omega_{\text{raw}}} \beta)^T \hat{\Sigma}_{\omega_{\text{raw}}} (\nabla_{\omega_{\text{raw}}} \beta) = \omega^T \hat{\Sigma}_{\omega_{\text{raw}}} \omega \quad (\text{A9})$$

Covariance of ω

The normalized weights are $\omega = g_2(\omega_{\text{raw}}) = \omega_{\text{raw}} / \|\omega_{\text{raw}}\|_2$.

$$\text{Cov}(\hat{\omega}) \approx \mathbf{J}_{\omega} \hat{\Sigma}_{\omega_{\text{raw}}} \mathbf{J}_{\omega}^T = \frac{1}{\beta^2} (I - \omega \omega^T) \hat{\Sigma}_{\omega_{\text{raw}}} (I - \omega \omega^T) \quad (\text{A10})$$

Appendix A.3. Observed Information Matrix

The asymptotic covariance matrix $\hat{\Sigma}_{\omega_{\text{raw}}}$ is estimated using the inverse of the observed information matrix. For the canonical ensemble model, the Hessian of the negative log-likelihood function is:

$$\mathbf{H} = \beta^2 \sum_n (\mathbf{X}_n^T \text{Diag}(P_n) \mathbf{X}_n - (\mathbf{X}_n^T P_n)(\mathbf{X}_n^T P_n)^T) \quad (\text{A11})$$

where \mathbf{X}_n is the matrix of attributes for choice set n and P_n is the vector of choice probabilities. The asymptotic covariance matrix is then given by:

$$\hat{\Sigma}_{\omega_{\text{raw}}} = \mathbf{H}^{-1} \quad (\text{A12})$$

Appendix A.4. Testing Temporal Stability of β

To assess whether the overall sensitivity to energy differences, represented by the scale parameter β , remains stable across the twelve-month study period, we conduct a formal hypothesis test. While we expect the normalized weights ω to show some variation due to changing conditions, we hypothesize that the fundamental choice sensitivity β is consistent.

Obtaining Standard Errors for β

First, we estimate the scale parameter $\hat{\beta}^{(t)}$ and its associated standard error, $SE(\hat{\beta}^{(t)})$, for each month $t \in \{1, \dots, 12\}$. Since β is a nonlinear transformation of the unconstrained parameters ω_{raw} , we use the Delta method to calculate its variance. The core idea of the Delta method is to use a first-order Taylor expansion to approximate a nonlinear function with a linear one, which allows the uncertainty of the original parameters to be translated into the uncertainty of the new, transformed parameters.

The variance of the scale parameter is calculated using the following formula:

$$\text{Var}(\hat{\beta}) \approx (\nabla_{\omega_{\text{raw}}} \beta)^T \hat{\Sigma}_{\omega_{\text{raw}}} (\nabla_{\omega_{\text{raw}}} \beta) \quad (\text{A13})$$

where $\nabla_{\omega_{\text{raw}}} \beta$ is the gradient of the scale parameter with respect to the raw parameter vector, and $\hat{\Sigma}_{\omega_{\text{raw}}}$ is the covariance matrix of the raw parameters. We approximate $\hat{\Sigma}_{\omega_{\text{raw}}}$ by inverting the Hessian matrix of the negative log-likelihood function,

$$\widehat{\text{Cov}}(\hat{\omega}_{\text{raw}}) = [-\mathbf{H}(\hat{\omega}_{\text{raw}})]^{-1}. \quad (\text{A14})$$

Appendix B. 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. The complete step-by-step derivations, including the full MLE equivalence proof, are provided in Supplementary Material Section 16.

Appendix B.1. Notation

Let O denote the set of all OD pairs, N_o the set of commuters in OD pair o , and R_o the number of route alternatives. For commuter $n \in N_o$ and route $j \in \{1, \dots, R_o\}$, let P_{jn} denote the choice probability, $\mathbf{X}_{jn} \in \mathbb{R}^M$ the attribute vector, and $y_{jn} \in \{0, 1\}$ the observed choice indicator.

Appendix B.2. Objective Function

We maximize the total Shannon entropy:

$$S = - \sum_{o \in O} \sum_{n \in N_o} \sum_{j=1}^{R_o} P_{jn} \ln P_{jn} \quad (\text{B1})$$

Appendix B.3. Model 1 (City-Wide Pooled Estimation)

This variant assumes spatial homogeneity: all commuters share a single parameter vector $\boldsymbol{\omega}_{\text{raw}} \in \mathbb{R}^M$. Maximizing Equation (B1) subject to (i) per-commuter normalization $\sum_j P_{jn} = 1$ and (ii) the global expected-value constraint

$$\sum_{o,n,j} P_{jn} \mathbf{X}_{jn} = \sum_{o,n,j} y_{jn} \mathbf{X}_{jn} \quad (\text{B2})$$

yields, via standard Lagrangian optimization (see Supplementary Section S16 for derivation):

$$P_{jn} = \frac{\exp(-\boldsymbol{\omega}_{\text{raw}}^\top \mathbf{X}_{jn})}{\sum_{j'=1}^{R_o} \exp(-\boldsymbol{\omega}_{\text{raw}}^\top \mathbf{X}_{j'n})} \quad (\text{B3})$$

where $\boldsymbol{\omega}_{\text{raw}}$ is the Lagrange multiplier enforcing Equation (B2).

Appendix B.4. MLE–MaxEnt Equivalence

Maximizing the log-likelihood $\ell = \sum_{o,n,j} y_{jn} \ln P_{jn}$ with P_{jn} given by Equation (B3) produces the first-order condition

$$\sum_{n=1}^N \sum_{j \in C_n} \mathbf{X}_{nj} [y_{nj} - P_{nj}(\boldsymbol{\omega}_{\text{raw}})] = 0 \quad (\text{B4})$$

which is algebraically identical to Equation (B2). The Lagrange multiplier and the MLE estimator therefore coincide, establishing the well-known duality between MaxEnt inference and discrete choice estimation [52,62].

Decomposing $\boldsymbol{\omega}_{\text{raw}} = \beta \boldsymbol{\omega}$ with $\beta = \|\boldsymbol{\omega}_{\text{raw}}\|_2$ and $\|\boldsymbol{\omega}\|_2 = 1$ gives:

$$P_{jn} = \frac{\exp(-\beta \boldsymbol{\omega}^\top \mathbf{X}_{jn})}{\sum_{j'=1}^{R_o} \exp(-\beta \boldsymbol{\omega}^\top \mathbf{X}_{j'n})} \quad (\text{B5})$$

Here $\boldsymbol{\omega}$ captures relative attribute preferences and β controls overall cost sensitivity. These parameters are global: they represent city-wide average preferences and cost sensitivity.

Appendix B.5. Model 2 (OD-Specific Estimation)

This variant assumes spatial heterogeneity: each OD pair o constitutes an independent sub-system with its own parameter vector $(\boldsymbol{\omega}_{\text{raw}})_o$. The normalization constraint is identical to Model 1. The expected-value constraint is imposed per OD pair:

$$\sum_{n \in N_o} \sum_{j=1}^{R_o} P_{jn}^o \mathbf{X}_{jn} = \sum_{n \in N_o} \sum_{j=1}^{R_o} y_{jn} \mathbf{X}_{jn}, \quad \forall o \in O \quad (\text{B6})$$

Because the sub-systems are independent, the total entropy decomposes additively and each OD pair can be optimized separately. The resulting choice probability takes the same MNL form as Equation (B3) with $(\boldsymbol{\omega}_{\text{raw}})_o$ replacing $\boldsymbol{\omega}_{\text{raw}}$, and the same MLE–MaxEnt equivalence holds within each OD pair. Decomposing $(\boldsymbol{\omega}_{\text{raw}})_o = \beta^o \boldsymbol{\omega}^o$ yields OD-specific parameters characterizing the cost sensitivity and attribute preferences of each sub-population.

References

1. Ardeshiri, A.; Vij, A. Lifestyles, Residential Location, and Transport Mode Use: A Hierarchical Latent Class Choice Model. *Transp. Res. Part A Policy Pract.* **2019**, *126*, 342–359.
2. Wang, H.; Yan, X.Y.; Wu, J. Free Utility Model for Explaining the Social Gravity Law. *J. Stat. Mech.* **2021**, *2021*, 033418.
3. Zhang, L.; Chen, T.; Liu, Z.; et al. Analysis of Multi-Modal Public Transportation System Performance Under Metro Disruptions: A Dynamic Resilience Assessment Framework. *Transp. Res. Part A Policy Pract.* **2024**, *183*, 104077.
4. Akamatsu, T.; Satsukawa, K.; Oyama, Y. Global Stability of Day-to-Day Dynamics for Schedule-Based Markovian Transit Assignment with Boarding Queues. *arXiv* **2023**, arXiv:2304.02194.
5. Deng, J.; Li, T.; Yang, Z.; et al. Heterogeneity in Route Choice During Peak Hours: Implications on Travel Demand Management. *Travel Behav. Soc.* **2025**, *38*, 100922.
6. Deng, J.; Yang, Y.; Li, T.; et al. Unveiling Route Choice Preferences and Classifying Travelers Based on Distinct Travel Patterns Using Trajectory Data. *Case Stud. Transp. Policy* **2025**, *19*, 101399.
7. Sakamane, P.; Phithakkitnukoon, S.; Smoreda, Z.; et al. Methods for Inferring Route Choice of Commuting Trip from Mobile Phone Network Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 306.
8. Kaneko, N.; Oka, H.; Chikaraishi, M.; et al. Route Choice Analysis in the Tokyo Metropolitan Area Using a Link-Based Recursive Logit Model Featuring Link Awareness. *Transp. Res. Procedia* **2018**, *34*, 251–258.
9. Li, D.; Miwa, T.; Morikawa, T.; et al. Incorporating Observed and Unobserved Heterogeneity in Route Choice Analysis with Sampled Choice Sets. *Transp. Res. Part C Emerg. Technol.* **2016**, *67*, 31–46.
10. Arriagada, J.; Prato, C.; Munizaga, M. Incorporating the Inertia Effect into a Route Choice Model Using Fare Transaction Data from a Large-Scale Public Transport Network. *Transp. Res. Part A Policy Pract.* **2025**, *196*, 104467.
11. Fosgerau, M.; Paulsen, M.; Rasmussen, T.K. A Perturbed Utility Route Choice Model. *Transp. Res. Part C Emerg. Technol.* **2022**, *136*, 103514.
12. Cazor, L.; Duncan, L.C.; Watling, D.P.; et al. A Closed-Form Bounded Route Choice Model Accounting for Heteroscedasticity, Overlap, and Choice Set Formation. *Transp. Res. Part B Methodol.* **2025**, *199*, 103275.
13. Wu, J.; Qu, Y.; Sun, H.; et al. Data-Driven Model for Passenger Route Choice in Urban Metro Network. *Physica A* **2019**, *524*, 787–798.
14. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*; Zarembka, P., Ed.; Academic Press: New York, NY, USA, 1972; pp. 105–142.
15. Train, K.E. *Discrete Choice Methods with Simulation*; Cambridge University Press: Cambridge, UK, 2009.
16. Ben-Akiva, M.E.; Lerman, S.R. *Discrete Choice Analysis: Theory and Application to Travel Demand*; MIT Press: Cambridge, MA, USA, 1985.
17. Ma, S.; Yu, Z.; Liu, C. Nested Logit Joint Model of Travel Mode and Travel Time Choice for Urban Commuting Trips in Xi'an, China. *J. Urban Plan. Dev.* **2020**, *146*, 04020020.
18. Garcia-Martinez, A.; Cascajo, R.; Jara-Diaz, S.R.; et al. Transfer Penalties in Multimodal Public Transport Networks. *Transp. Res. Part A Policy Pract.* **2018**, *114*, 52–66.
19. Okubo, T.; Kitano, N.; Morimoto, A. A Transportation Choice Model on the Commuter Railroads Using Inverse Reinforcement Learning. *Asian Transp. Stud.* **2022**, *8*, 100072.
20. McFadden, D.; Train, K. Mixed MNL Models for Discrete Response. *J. Appl. Econom.* **2000**, *15*, 447–470.
21. Greene, W.H.; Hensher, D.A. A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. *Transp. Res. Part B Methodol.* **2003**, *37*, 681–698.
22. Kim, S.H.; Mokhtarian, P.L. Finite Mixture (or Latent Class) Modeling in Transportation: Trends, Usage, Potential, and Future Directions. *Transp. Res. Part B Methodol.* **2023**, *172*, 134–173.
23. Ghorbani, A.; Nassir, N.; Lavieri, P.S.; et al. Enhanced Utility Estimation Algorithm for Discrete Choice Models in Travel Demand Forecasting. *Transportation* **2025**. <https://doi.org/10.1007/s11116-024-10579-1>.
24. Tang, Y.; Jiang, Z.; Zou, X.; et al. Mining Motif Periodic Frequent Travel Patterns of Individual Metro Passengers Considering Uncertain Disturbances. *Int. J. Transp. Sci. Technol.* **2024**, *15*, 102–121.
25. Mohammed, M.; Oke, J. Origin-Destination Inference in Public Transportation Systems: A Comprehensive Review. *Int. J. Transp. Sci. Technol.* **2023**, *12*, 315–328.
26. Yap, M.; Wong, H.; Cats, O. Public Transport Crowding Valuation in a Post-Pandemic Era. *Transportation* **2025**, *52*, 287–306.
27. Cherchi, E.; De Dios Ortúzar, J. Empirical Identification in the Mixed Logit Model: Analysing the Effect of Data Richness. *Netw. Spat. Econ.* **2008**, *8*, 109–124.
28. Sadeghian, P.; Zhao, X.; Golshan, A.; et al. A Stepwise Methodology for Transport Mode Detection in GPS Tracking Data. *Travel Behav. Soc.* **2022**, *26*, 159–167.
29. Montini, L.; Antoniou, C.; Axhausen, K.W. Route and Mode Choice Models Using GPS Data. Available online: <https://ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/reports/1201-1300/ab1204.pdf> (accessed on 15 September 2025).
30. Helbing, D.; Nagel, K. The Physics of Traffic and Regional Development. *Contemp. Phys.* **2004**, *45*, 405–426.

31. Ozaki, J.; Shida, Y.; Takayasu, H.; et al. Direct Modelling from GPS Data Reveals Daily-Activity-Dependency of Effective Reproduction Number in COVID-19 Pandemic. *Sci. Rep.* **2022**, *12*, 17888.
32. Shida, Y.; Ozaki, J.; Takayasu, H.; et al. Potential Fields and Fluctuation-Dissipation Relations Derived from Human Flow in Urban Areas Modeled by a Network of Electric Circuits. *Sci. Rep.* **2022**, *12*, 9918.
33. Ministry of Land, Infrastructure, Transport and Tourism (MLIT). National Land Numerical Information: Railway Data (N02). Available online: <https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-N02-2023.html> (accessed on 14 October 2025). (In Japanese)
34. Su, R.; McBride, E.C.; Goulias, K.G. Pattern Recognition of Daily Activity Patterns Using Human Mobility Motifs and Sequence Analysis. *Transp. Res. Part C Emerg. Technol.* **2020**, *120*, 102796.
35. Zhang, X.; Li, N. An Activity Space-Based Gravity Model for Intracity Human Mobility Flows. *Sustain. Cities Soc.* **2024**, *101*, 105073.
36. Ministry of Land, Infrastructure, Transport and Tourism (MLIT). Transfer Survey Report 2016. Available online: <https://www.mlit.go.jp/common/001179761.pdf> (accessed on 15 September 2025). (In Japanese)
37. Takahashi, T. Transportation Mode Choice and Spatial Structure of a City. *Technical Report, Center for Spatial Information Science*; Working Paper; The University of Tokyo: Tokyo, Japan, 2019. Available online: <https://www.csis.u-tokyo.ac.jp/en/research/> (accessed on 15 September 2025). (In Japanese)
38. Sadeghian, P.; Håkansson, J.; Zhao, X. Review and Evaluation of Methods in Transport Mode Detection Based on GPS Tracking Data. *J. Traffic Transp. Eng. (Engl. Ed.)* **2021**, *8*, 467–482.
39. Markos, C.; Yu, J.J. Unsupervised Deep Learning for GPS-Based Transportation Mode Identification. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020.
40. Montazeri-Gh, M.; Fotouhi, A. Traffic Condition Recognition Using the K-Means Clustering Method. *Sci. Iran.* **2011**, *18*, 930–937.
41. Rasmussen, T.K.; Ingvarsson, J.B.; Halldórsdóttir, K.; et al. Improved Methods to Deduct Trip Legs and Mode from Travel Surveys Using Wearable GPS Devices: A Case Study from the Greater Copenhagen Area. *Comput. Environ. Urban Syst.* **2015**, *54*, 301–313.
42. Zheng, Y.; Liu, L.; Wang, L.; et al. Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web. In Proceedings of the 17th International Conference on World Wide Web, Beijing, China, 21–25 April 2008; pp. 247–256.
43. Dabiri, S.; Heaslip, K. Inferring Transportation Modes from GPS Trajectories Using a Convolutional Neural Network. *Transp. Res. Part C Emerg. Technol.* **2018**, *86*, 360–371.
44. Zhu, Q.; Zhu, M.; Li, M.; et al. Identifying Transportation Modes from Raw GPS Data. In *Social Computing*; Che, W., Han, Q., Wang, H., et al., Eds.; Springer: Singapore, 2016; Volume 623, pp. 395–409.
45. Tokyo Metropolitan Government. Current Status and Issues of Regional Public Transportation. Available online: https://www.toshiseibi.metro.tokyo.lg.jp/documents/d/toshiseibi/pdf_bunyabetsu_kotsu_butsuryu_pdf_chiiki_koutsu_kentoukai01_2-1 (accessed on 19 June 2026). (In Japanese)
46. Schneider, C.M.; Belik, V.; Couronné, T.; et al. Unravelling Daily Human Mobility Motifs. *J. R. Soc. Interface* **2013**, *10*, 20130246.
47. Jiang, S.; Ferreira, J.; Gonzalez, M.C. Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE Trans. Big Data* **2017**, *3*, 208–219.
48. Cao, J.; Li, Q.; Tu, W.; et al. Characterizing Preferred Motif Choices and Distance Impacts. *PLoS One* **2019**, *14*, e0215242.
49. At Home Co., Ltd. Trend Survey-Tokyo. Available online: <https://www.athome.co.jp/corporate/wp-content/themes/news/pdf/densha-tsuukin-201806/densha-tsuukin-201806.pdf> (accessed on 3 March 2026). (In Japanese)
50. Hibino, N.; Morita, Y.; Uchiyama, H. A Study on Method of Setting Alternative Routes for Analysis on Route Choice Behavior. *J. City Plan. Inst. JPN* **2001**, *36*, 787–792. (In Japanese)
51. Dixit, M.; Cats, O.; Brands, T.; et al. Perception of Overlap in Multi-Modal Urban Transit Route Choice. *Transp. A Transp. Sci.* **2023**, *19*, 2005180.
52. Anas, A. Discrete Choice Theory, Information Theory and the Multinomial Logit and Gravity Models. *Transp. Res. Part B Methodol.* **1983**, *17*, 13–23.
53. Yoo, S.; Kumagai, J.; Aki, R.; et al. Railway Network Expansion Reduces Air Pollution in Tokyo over 25 Years. *Sustain. Cities Soc.* **2025**, *127*, 106408.
54. Neter, J.; Kutner, M.H.; Nachtsheim, C.J.; et al. *Applied Linear Statistical Models*; Irwin: Chicago, IL, USA, 1996.
55. Hair, J.F.; Black, W.C.; Babin, B.J.; et al. *Multivariate Data Analysis*, 7th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2009.
56. Casella, G.; Berger, R.L. *Statistical Inference*, 2nd ed.; Duxbury Press: Pacific Grove, CA, USA, 2002.
57. Greene, W.H. *Econometric Analysis*, 8th ed.; Pearson Education: New York, NY, USA, 2018.
58. Ozili, P.K. The Acceptable R-Square in Empirical Modelling for Social Science Research **2023**, 134–143. <https://doi.org/10.2139/ssrn.4128165>.

59. Ji, Y.; Jiang, S.; Du, Y.; et al. Estimation of Bimodal Urban Link Travel Time Distribution and Its Applications in Traffic Analysis. *Math. Probl. Eng.* **2015**, 2015, 615468.
60. Kumagai, J. Visualizing Railway Transfer Penalties and Their Effects on Population Distribution in the Tokyo Metropolitan Area. *Future Transp.* **2025**, 5, 114.
61. Wilson, A. A Statistical Theory of Spatial Distribution Models. *Transp. Res.* **1967**, 1, 253–269.
62. Donoso, P.; de Grange, L. A Microeconomic Interpretation of the Maximum Entropy Estimator of Multinomial Logit Models and Its Equivalence to the Maximum Likelihood Estimator. *Entropy* **2010**, 12, 2077–2084.
63. Ministry of Land, Infrastructure, Transport and Tourism (MLIT). Transfer Survey Report 2017. Available online: <https://www.mlit.go.jp/common/001178977.pdf> (accessed on 5 April 2025). (In Japanese)
64. Vittinghoff, E.; McCulloch, C.E. Relaxing the Rule of Ten Events Per Variable in Logistic and Cox Regression. *Am. J. Epidemiol.* **2007**, 165, 710–718.