Multivariate Discrete Choice with Rational Inattention: Model Development, Application, and Calibration

Abstract

The recent application of the rational inattention (RI) theory in transportation has shed light on a promising alternative way of understanding how information influences the travel choices of passengers. However, existing RI literature has not yet addressed the discrete choice problem with multiple variates. Thus, this study develops a multivariate rational inattention (MRI) discrete choice model. This assumes that acquiring information is costly and the unit information cost varies among variates, so decision-makers rationally choose the amount of information to acquire for each variate. We demonstrate that the MRI discrete choice model results in a probabilistic formulation similar to the logit model, but with the superiority of integrating unit information costs and the prior knowledge of decision-makers. Furthermore, we apply the MRI discrete choice model to the metro route choice problem and calibrate the model based on the revealed preference (RP) data collected from the Chengdu metro. It is found that the proposed model has satisfactory accuracy with better interpretability than the logit model and univariate rational inattention discrete choice model.

Keywords

Multivariate Discrete Choice; Rational Inattention; Model Calibration; Route Choice

1 Introduction

Understanding and modeling the route choice preferences of passengers is essential to refining supply decisions that affect public transport planning and operations (Amirgholy et al., 2017, Huang et al., 2016, Tirachini et al., 2010). It is also the foundation for the transit assignment problem and the valuation of transit attributes such as time and crowding (Björklund and Swärdh, 2017, Wardman and Whelan, 2011).

Advancements in technology, including smartphone applications, mapping software, and social media, have significantly reshaped route choice preferences by providing multiple information sources to passengers. The role of information in modeling route choice preferences has thus gained significant research interest (Ben-Elia and Avineri, 2015; Chorus et al., 2006, De Palma et al., 2012). This particularly underscores the influence of information acquisition on passenger perceptions. A variety of research questions have been proposed, including the estimation of information confidence intervals (Ettema and Timmermans, 2006), the optimization of information releasing rates (Yin et al., 2019), the categorization of information types (De Palma et al., 2012), and the classification of traveler sensitivities to information provision (Zhu et al., 2019).

Considering the variety of information sources, full comprehension of the information that passengers acquire and the impact of this information on passenger perceptions is almost impossible. The finite capacity of human information processing leads decision-makers (DM) to consider information selectively (Hogarth and Wiley, 1980). This implies that the acquisition of all available information is unlikely for DMs. In such a context, the theory of rational inattention (RI) (Maćkowiak et al., 2020, Sims, 2003, 2010) offers a novel perspective to model the mechanism of information selective behavior by considering the cost of information acquisition within the framework of utility-maximizing behaviors of DMs, wherein the costs may arise from time or cognitive efforts. This theory has been applied across various domains, such as a linear-quadratic-Gaussian control setup (Miao et al., 2019), static finance models (Kacperczyk et al., 2016), and discrete choice preference analysis (Matějka and McKay, 2015).

According to a review by Maćkowiak et al. (2020), previous RI studies have predominantly addressed discrete choices featuring only a single influence variate. Matějka and McKay (2015), henceforth referred to as MM, initially proposed the univariate RI (URI) discrete choice model, drawing upon the difference between unconditional and conditional choice probability to ascertain the amount of information that DMs acquire. Following their work, Caplin et al. (2019) introduced an iterative method to calculate the unconditional choice probability and derived necessary and sufficient conditions to solve the URI discrete choice model. Fosgerau et al. (2020) extended the previous model by providing a general equivalence between the additive random utility discrete choice and RI-based models.

Nevertheless, discrete choices (for example, route choices) in the transportation domain are commonly affected by a multivariate situation, where each variate potentially presents disparate information acquisition difficulty and affects choice quality. Consequently, DMs may exhibit varying preferences regarding information pertinent to different variates, which the URI discrete model by MM fails to capture. Recently, Habib (2023) extended and calibrated the URI discrete model for the multivariate influenced commuting mode choice problem. However, Habib (2023) did not distinguish the DM's preference difference among variates and used actual market share to measure the unconditional choice probability, which may not always be available.

Prior research into the DMs' varying information preferences has been informed by studies on attribute non-attention (ANA) (Hensher, 2014, Hensher and Greene, 2010, Hensher and Rose, 2009). These studies drew on data from stated preference surveys and offered the DMs all information about the choice directly. Therefore, the studies did not involve the information-seeking process in practice. Nevertheless, they showed that the DMs could disregard certain variates during decision-making. This finding inspired our investigation into how information from various sources shapes the preferences of the DMs for different variates during the discrete decision-making processes in actual practice, such as when choosing a travel route.

To model the aforementioned information preference among different variates, we formulate a multivariate (MRI) discrete choice model by factoring in the distinct information costs with different variates. Furthermore, we calibrate the developed MRI discrete choice model with metro data to understand the route choice preferences of travelers. To be more specific, we utilize the revealed preference (RP) data from the Chengdu metro, a collection informed by inferred passenger time-space trajectories from our earlier work (Chen et al., 2023). The findings from our research not only underscore the significant variations in information preference among different variates but also demonstrate the superiority of the MRI discrete choice model over the conventional generalized multinomial logit (MNL) model, offering enhanced interpretability and maintaining an equivalent level of predictive fidelity.

To summarize, the contributions of this study are threefold. 1) The multivariate discrete choice is modeled with rational inattention theory, which can capture the influence of information on different variates without assuming the source or content of information. 2) The proposed model is applied to the metro route choice problem, revealing the behavioral patterns of metro passengers with imperfect information. 3) The rational inattention theory is calibrated with a real case study.

The remainder of this paper is structured as follows. Section 2 revisits the URI discrete choice model. Section 3 presents the modeling of the MRI discrete choice problem. Section 4 presents the applications of the MRI discrete choice model for the metro route choice problem and develops the methodology for the calibration of the model. Section 5 discusses the result of the real-case study using Chengdu Metro data. Finally, Section 5.2.4 discusses and concludes the study.

2 Revisiting the Univariate Rational Inattention Discrete Choice Model

Before introducing the MRI discrete choice problem, this section revisits the URI discrete choice model with MM as a prerequisite. The section commences with a conceptual review of the URI discrete choice problem, followed by the modeling method and solution method. Interested readers are referred to the full paper for a complete description (Matějka and McKay, 2015).

2.1 The URI discrete choice problem

The URI discrete choice problem describes a scenario where a DM selects from a set of alternatives $\{\rho = 1, 2, \ldots\}$. Each alternative ρ can offer a DM a unique utility payoff v_ρ. Consequently, the utility payoffs of the alternative set form a vector $\mathbf{v} = \begin{bmatrix} v_1 & \cdots & v_\rho & \cdots \end{bmatrix}^T$. The DM's objective is to choose the alternative with the highest utility payoff.

The state of $\mathbf{v}$ belongs to a finite set with a corresponding probability distribution. Correspondingly, the DM is not sure about the true state of $\mathbf{v}$ but does possess knowledge about the distribution of $\mathbf{v}$. This knowledge is named prior knowledge, is denoted by $p(\mathbf{v})$, is derived from the DM's experience, and remains consistent regardless of the realized state for a specific choice. However, this prior knowledge may not be sufficient for identifying the alternative with the highest utility payoff for each state.

To mitigate this uncertainty, the DM can seek additional information, defined as signals, to update their prior knowledge to posterior knowledge. Benefiting from the information provided by signals, the posterior knowledge is more exact than the prior knowledge. A signal, denoted as a vector $\mathbf{s}$, is a random variable that carries information about the state. For convenience, see the study of MM (Matějka and McKay, 2015) for the details of this signal.

However, seeking information incurs costs, with more informative signals generally being more expensive in terms of time and cognitive effort. Conversely, a more informative signal leads to more precise knowledge about the state of the utility payoff vector $\mathbf{v}$, aiding in decision-making. Consequently, the DM must balance the cost of information against the benefits of an informed decision by determining a preference among signals. The URI discrete choice framework is depicted below.

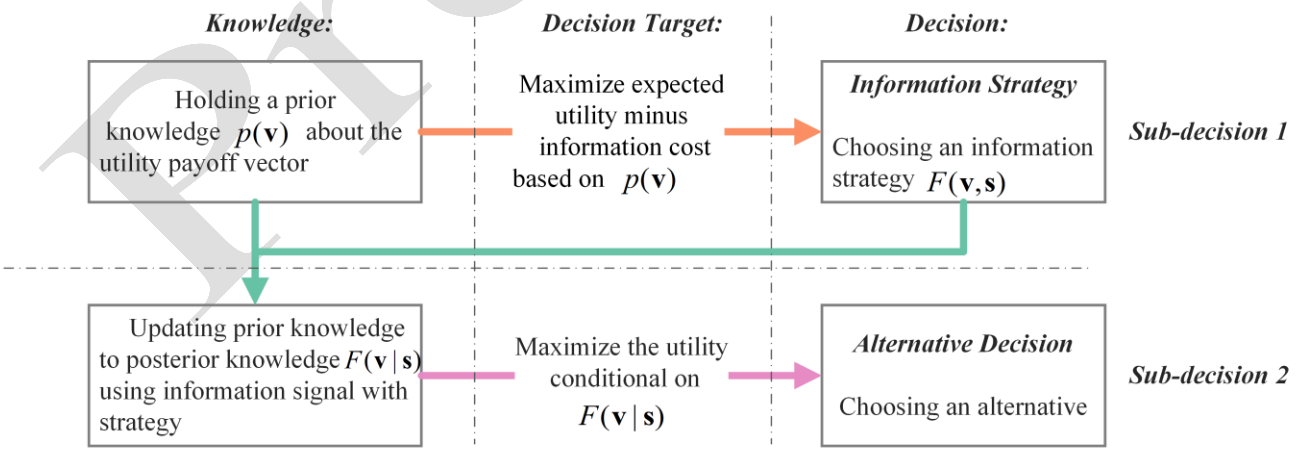


Fig. 1 Decision process for URI discrete choice problem

As depicted in Fig. 1, the MM divided the URI discrete choice decision process into two sub-decisions: The decision of information strategy and the decision of alternatives.

i. The Decision of Information Strategy: As indicated by the orange arrow, the DM determines their preference among signals based on prior knowledge. This is referred to as the information strategy, which is a joint distribution of utility payoff and signals, denoted by $F(\mathbf{v},\mathbf{s})$.

ii. The Decision of Alternatives: As indicated by the green arrow, the DM seeks information for the chosen strategy and updates the prior knowledge to the posterior knowledge. As indicated by the purple arrow, the DM makes a decision based on the posterior knowledge.

The key elements in these two sub-decisions, i.e., prior knowledge, information strategy, and posterior knowledge, are highly interrelated. The marginal distribution of the information strategy equals the prior knowledge, as shown in Eq. (1), ensuring that the DM's posterior knowledge is consistent with their prior knowledge.

$$\int F(\mathbf{v},\mathbf{s})d\mathbf{s} = p(\mathbf{v}). \tag{1}$$

Given this mathematical hold, the DM chooses the probabilistic characteristics of the conditional distribution $F(\mathbf{s}\,|\,\mathbf{v})$ in the information decision. In contrast, $F(\mathbf{v}\,|\,\mathbf{s})$ is the posterior knowledge after receiving signals.

Given the above-defined relations, the decision of the first sub-decision influences that of the second sub-decision. The decision target of the first sub-decision should account for that of the second sub-decision accordingly. The goals of this are twofold: To maximize the ex ante expected utility payoff and to minimize the information cost. The ex ante expected utility payoff measures the indirect impact of the information strategy on the alternative decision target as it influences the posterior knowledge. The decision target of the second sub-decision is intended to maximize the expected utility payoff according to the posterior knowledge. The modeling of the decision process is as follows.

i. This study denotes the maximized expected utility payoff corresponding to a posterior knowledge in the alternative decision as $A(\cdot)$ and denotes the decision result as $a(\cdot)$, which can be calculated as follows:

$$A\left(F(\mathbf{v}\,|\,\mathbf{s})\right) = \max_{\forall \rho} E\left(v_\rho \,|\, F(\mathbf{v}\,|\,\mathbf{s})\right), \tag{2}$$

$$a\left(F\left(\mathbf{v}\,|\,\mathbf{s}\right)\right)=\underset{\forall\rho}{\operatorname{argmax}}\,E\left(v_{\rho}\,|\,F\left(\mathbf{v}\,|\,\mathbf{s}\right)\right). \tag{3}$$

ii. This study denotes the ex ante expected utility payoff corresponding to an information strategy as a function $U(\cdot)$, which can be calculated as follows:

$$U\left(F\left(\mathbf{v},\mathbf{s}\right)\right)=\int\left(\int A\left(F\left(\mathbf{v}\,|\,\mathbf{s}\right)\right)F\left(\mathbf{s}\,|\,\mathbf{v}\right)d\mathbf{s}\right)p\left(\mathbf{v}\right)d\mathbf{v}. \tag{4}$$

iii. The information cost for an information strategy decision is the product of the unit information cost λ and the difference (mutual information) between the Shannon entropy of prior knowledge and the expected Shannon entropy of posterior knowledge. This study denotes the Shannon entropy function as $H(\cdot)$, the expected Shannon entropy of posterior knowledge corresponding to an information strategy as $E_H(\cdot)$, and the information cost function as $\hat{c}(\cdot)$, which can be calculated as follows:

$$H(f)=-\int f(x)\log\left(f(x)\right)dx, \tag{5}$$

$$E_H\left(F\left(\mathbf{v},\mathbf{s}\right)\right)=\int\left(\int H\left(F\left(\mathbf{v}\,|\,\mathbf{s}\right)\right)F\left(\mathbf{s}\,|\,\mathbf{v}'\right)d\mathbf{s}\right)p\left(\mathbf{v}'\right)d\mathbf{v}', \tag{6}$$

$$\hat{c}\left(F\left(\mathbf{v},\mathbf{s}\right)\right)=\lambda\left(H\left(p(\mathbf{v})\right)-E_H\left(F\left(\mathbf{v},\mathbf{s}\right)\right)\right). \tag{7}$$

iv. The decision target of the information strategy can then be formulated as follows:

$$I\left(F\left(\mathbf{v},\mathbf{s}\right)\right)=\max_{F(\mathbf{v},\mathbf{s})}\left(U\left(F\left(\mathbf{v},\mathbf{s}\right)\right)-\hat{c}\left(F\left(\mathbf{v},\mathbf{s}\right)\right)\right). \tag{8}$$

2.2 The solution method of the URI discrete choice model

The goal for solving the URI discrete choice model is the derivation of the probability of the DM choosing each alternative that is conditional on the utility vector (denoted by $P(\rho\,|\,\mathbf{v})$). Considering the diversity of information, solving Eqs. (4)–(8) by enumerating the signal and posterior knowledge is impractical for real-life problems. MM demonstrated that solving these equations could be accomplished without explicitly investigating signals and posterior knowledge. This subsection briefly reviews MM's method.

MM handled the signal and the posterior knowledge implicitly by proving a lemma: *Each* *alternative is selected in at most one posterior knowledge under an optimal information strategy*, which is indexed as LEMMA 1 by MM. LEMMA 1 indicates that a certain alternative decision only corresponds to one unique posterior knowledge. Accordingly, MM derived equivalent transformations for Eq. (4) and Eq. (7) when the information strategy was optimal. Specifically,

i. Representing Eq. (4) with the choice probabilities $P(\rho \mid \mathbf{v})$ conditional on the utility payoff vector $\mathbf{v}$ and prior knowledge $p(\mathbf{v})$.

$$\begin{aligned} U\left(F\left(\mathbf{v},\mathbf{s}\right)\right) &= \int\left(\int A\left(F\left(\cdot \mid \mathbf{s}\right)\right) F(\mathbf{s} \mid \mathbf{v}) d\mathbf{s}\right) p\left(\mathbf{v}\right) d\mathbf{v} \\ &= \sum_{\forall \rho} \int v_\rho P\left(\rho \mid \mathbf{v}\right) p\left(\mathbf{v}\right) d\mathbf{v} \end{aligned} . \tag{9}$$

ii. Representing Eq. (7) with the mutual information for the unconditional probabilities $P(\rho)$ and conditional probabilities $P(\rho \mid \mathbf{v})$. The unconditional probabilities are the integral of the conditional probabilities over the utility payoff vector $\mathbf{v}$.

$$\begin{aligned} \hat{c}\left(F\left(\mathbf{v},\mathbf{s}\right)\right) &= \lambda\left(H\left(p(\mathbf{v})\right) - H\left(F\left(\mathbf{v} \mid \mathbf{s}\right)\right)\right) \\ &= \lambda\left(-H\left(P\left(\rho\right)\right) + \int H\left(P\left(\rho \mid \mathbf{v}\right)\right) p\left(\mathbf{v}\right) d\mathbf{v}\right) \end{aligned} . \tag{10}$$

Based on the transformations above, MM solved the analytical solution for the choice probabilities conditional on the utility payoff vector $\mathbf{v}$ as:

$$P\left(\rho \mid \mathbf{v}\right) = \frac{P\left(\rho\right) \exp\left(v_\rho / \lambda\right)}{\sum_{\forall \rho'} \exp\left(v_{\rho'} / \lambda\right)} . \tag{11}$$

Eq. (11) does not give a fully explicit expression for the choice probabilities because it depends on unconditional choice probabilities $P(\rho)$. To solve for $P(\rho)$, Caplin et al. (2019) developed an iteratively updated method based on LEMMA 2 proved by MM: *The collection of* $P(\rho \mid \mathbf{v})$ $\rho = 1, 2, \cdots$ *satisfies the optimal information if and only if the following formulation holds*.

$$\int \frac{\exp\left(v_\rho / \lambda\right)}{\sum_{\forall \rho'} P\left(\rho'\right) \exp\left(v_{\rho'} / \lambda\right)} p(\mathbf{v}) d\mathbf{v} = 1 . \tag{12}$$

In summary, the key points of the URI discrete choice model created by MM can be stated as follows: 1) The model considers the influence of information-seeking behavior on the DM's decision result by considering the information cost. 2) The information cost is modeled by the unit information cost and the mutual information between the DM's knowledge about the problem before and after seeking information, i.e., prior knowledge and posterior knowledge. 3) The mutual information between the prior and the posterior knowledge is represented by that between the DM's unconditional and conditional choice probability distribution.

3 Modeling Multivariate Rational Inattention Discrete Choice

This section explores the MRI discrete choice problem and extends the methodology established in the URI to formulate choice probabilities for the MRI context. The MRI discrete choice problem extends from the URI by incorporating multiple variates that influence decision-making. In contrast to the modeling of URI, the modeling of the MRI discrete choice problem should consider the information cost corresponding to each variate. Correspondingly, the method to represent the amount of information that the DM acquires for each method should be developed to solve the MRI discrete choice model.

3.1 Problem description

In the MRI discrete choice problem, the DM must select from a set of alternatives, for which the utility payoff of each alternative is specified by a set of variates. The utility vector $\mathbf{v}$ is a linear function of the variates with the coefficient vector $\boldsymbol{\beta} = \begin{bmatrix} \beta^1 & \cdots & \beta^\chi & \cdots \end{bmatrix}^T$. Letting $\mathbf{x}^\chi = \begin{bmatrix} x_1^\chi & \dots & x_\rho^\chi & \dots \end{bmatrix}^T$ be the vector of χ-th variate that specific the utility vector $\mathbf{v}$, and denoting $\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{x}^1 & \dots & \mathbf{x}^\chi & \dots \end{bmatrix}$, the utility vector $\mathbf{v}$ can then be calculated as $\mathbf{v} = \hat{\mathbf{x}}\boldsymbol{\beta}$.

To identify the utility payoff of each alternative, the DM must determine the state of the variate vectors. Similar to the URI discrete choice problem, the DM is not sure of the true state of any variate vector $\mathbf{x}^\chi$ but does have prior knowledge about variate vectors. Additionally, the DM can seek information to update these prior knowledge distributions. The information sources and the content for different variates may vary, leading to different levels of difficulty in acquiring information among variates. Additionally, each variate may affect the utility value differently. Given these facts, the DM may exhibit varying preferences regarding information about different variates. To capture these varying preferences, this study distinguishes the signal, unit information cost, information strategy, prior knowledge, and posterior knowledge for each variate in the modeling. The symbols of these elements in the MRI discrete choice problem are adapted from those of the URI, as follows.

i. Signal: Represents the signal about $\mathbf{x}^{\chi}$ and $\hat{\mathbf{x}}$ with $\mathbf{s}^{\chi}$ and $\hat{\mathbf{s}}$.

ii. Unit information cost: Represents the unit information cost corresponding to the χ-th variate with λ^{χ}.

iii. Prior knowledge: Represents the prior knowledge about $\mathbf{x}^{\chi}$ and $\hat{\mathbf{x}}$ with $p\left(\mathbf{x}^{\chi}\right)$ and $p\left(\hat{\mathbf{x}}\right)$. $p\left(\mathbf{x}^{\chi}\right)$ is the marginal distribution of $p\left(\hat{\mathbf{x}}\right)$ about $\mathbf{x}^{\chi}$.

iv. Information strategy: Represents the information strategy about $\mathbf{s}^{\chi}$ and $\hat{\mathbf{s}}$ with $F\left(\mathbf{x}^{\chi},\mathbf{s}^{\chi}\right)$ and $F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)$. $F\left(\mathbf{x}^{\chi},\mathbf{s}^{\chi}\right)$ is the marginal distribution of $F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)$ about $\left[\mathbf{x}^{\chi} \quad \mathbf{s}^{\chi}\right]$.

v. Posterior knowledge: Represents the posterior knowledge about $\mathbf{x}^{\chi}$ and $\hat{\mathbf{x}}$ with $F\left(\mathbf{x}^{\chi} \mid \mathbf{s}^{\chi}\right)$ and $F\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}\right)$.

In the context of metro transportation, a DM signifies a metro passenger, with the alternatives set embodying the routes available to choose from. x_{ρ}^{χ} can represent the variates corresponding to the passenger's ρ-th available route, such as the number of transfer times, the length of waiting and sitting time, and the level of crowding. To illustrate the MRI discrete choice concept, we present a simplified scenario in which a passenger must choose between two routes. The utility of each route is determined by two variates, which can represent the factors mentioned. Each variate has two equally probable states, as shown in Table 1. The passenger is aware of these states and their possibilities, representing prior knowledge. An in-depth discussion of the metro route choice problem is covered in Section 4.

Table 1 Illustration of the choice situation

Alternative	Alternative-1		Alternative-2	
Variate	x_1^1	x_1^2	x_2^1	x_2^2
State 1/Possibility	20/0.5	10/0.5	10/0.5	20/0.5
State 2/Possibility	50/0.5	20/0.5	20/0.5	55/0.5

Given that each variate has two states, $\mathbf{x}^1=\left[x_1^1 \quad x_2^1\right]^T$ and $\mathbf{x}^2=\left[x_1^2 \quad x_2^2\right]^T$ together have four combined states. This results in 16 possible states for $\hat{\mathbf{x}}$ when considering two routes, as shown in Table 2.

Table 2 Enumeration of the possible state of $\mathbf{x}^1$, $\mathbf{x}^2$, and $\hat{\mathbf{x}}$

$\hat{\mathbf{x}}$		$\mathbf{x}^1$			
		$[20 \quad 10]^T$	$[20 \quad 20]^T$	$[50 \quad 10]^T$	$[50 \quad 20]^T$
$\mathbf{x}^2$	$\begin{bmatrix}20\\55\end{bmatrix}$	$\begin{bmatrix}20 & 20\\10 & 55\end{bmatrix}$	$\begin{bmatrix}20 & 20\\20 & 55\end{bmatrix}$	$\begin{bmatrix}50 & 20\\10 & 55\end{bmatrix}$	$\begin{bmatrix}50 & 20\\20 & 55\end{bmatrix}$
	$\begin{bmatrix}20\\20\end{bmatrix}$	$\begin{bmatrix}20 & 20\\10 & 20\end{bmatrix}$	$\begin{bmatrix}20 & 20\\20 & 20\end{bmatrix}$	$\begin{bmatrix}50 & 20\\10 & 20\end{bmatrix}$	$\begin{bmatrix}50 & 20\\20 & 20\end{bmatrix}$
	$\begin{bmatrix}10\\55\end{bmatrix}$	$\begin{bmatrix}20 & 10\\10 & 55\end{bmatrix}$	$\begin{bmatrix}20 & 10\\20 & 55\end{bmatrix}$	$\begin{bmatrix}50 & 10\\10 & 55\end{bmatrix}$	$\begin{bmatrix}50 & 10\\20 & 55\end{bmatrix}$
	$\begin{bmatrix}10\\20\end{bmatrix}$	$\begin{bmatrix}20 & 10\\10 & 20\end{bmatrix}$	$\begin{bmatrix}20 & 10\\20 & 20\end{bmatrix}$	$\begin{bmatrix}50 & 10\\10 & 20\end{bmatrix}$	$\begin{bmatrix}50 & 10\\20 & 20\end{bmatrix}$

3.2 Modeling of the MRI discrete choice problem

Similar to the URI discrete choice problem, the decision of the DM in the MRI discrete choice problem has two stages: the information strategy decision and the alternative decision. The difference is that the DM should determine the optimal amount of information to acquire for each variate to maximize the expected utility while minimizing costs in the information strategy. The modeling of the decision process for the MRI discrete choice problem is updated from that of URI by redefining the ex ante expected utility payoff function and the information cost function.

i. The ex ante expected utility payoff function:

$$U\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)\right)=\int\left(\int A\left(F\left(\cdot\mid\hat{\mathbf{s}}\right)\right)F\left(\hat{\mathbf{s}}\mid\hat{\mathbf{x}}\right)d\hat{\mathbf{s}}\right)p\left(\hat{\mathbf{x}}\right)d\hat{\mathbf{x}}\,. \tag{13}$$

ii. The information cost function:

$$\hat{c}\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)\right)=\sum_{\forall\chi}\lambda^{\chi}\left(H\left(p(\mathbf{x}^{\chi})\right)-E_H\left(F\left(\mathbf{x}^{\chi},\mathbf{s}^{\chi}\right)\right)\right). \tag{14}$$

3.3 Solving the choice probability to the MRI discrete choice problem

This subsection presents the expansion and modification of the URI discrete choice model to resolve the MRI discrete choice problem.

The primary distinction between the URI and MRI discrete choice models is the fact that the MRI discrete choice model considers the amount of information obtained by the DM for each variate individually. This introduces a complexity that is not faced by the URI discrete choice model. In the URI discrete choice model, the key to deriving the choice probability formulation is solving for the information cost. This involves transforming the amount of information that the DM acquires for a univariate scenario into mutual information between the conditional and unconditional choice probabilities. However, the approach of the URI discrete choice model cannot be directly applied to the MRI discrete choice model. The challenge lies in 1) Proving that the primary lemmas still hold in the MRI discrete choice problem and 2) The fact that mutual information between the conditional and unconditional choice probabilities is insufficient to quantitatively represent the information that the DM acquires for multiple variates. The MRI discrete choice model must consider the interaction and the cumulative effect of information across multiple variates, making the problem significantly more complex.

3.3.1 *Information decision target transformation*

This study proves that LEMMA 1 by MM is still valid in the MRI discrete choice problem, which lays the groundwork for implicitly handling the signal and posterior knowledge. Specifically, by defining $\mathrm{S}_\rho = \left\{\hat{\mathbf{s}} : a\left(F\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}\right)\right) = \rho\right\}$ as the set of signals that lead to the DM choosing alternative ρ, the probability of the DM choosing the alternative ρ conditional to $\hat{\mathbf{x}}$ can be denoted by:

$$P\left(\rho \mid \hat{\mathbf{x}}\right) = \int_{\hat{\mathbf{s}} \in \mathrm{S}_\rho} F\left(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}\right) d\hat{\mathbf{s}}. \tag{15}$$

Then, the MRI version of LEMMA 1 can be described as follows.

LEMMA 1: If $F(\cdot)$ is the optimal information strategy for the DM, then the posterior knowledge $F\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}\right)$ are the same for all $\hat{\mathbf{s}} \in \mathrm{S}_\rho$.

Proof. See Appendix B.

Based on LEMMA 1, the Eq. (13) can be transferred as follows when $F(\cdot)$ is optimal. The derivation is outlined in Appendix C.

$$\begin{aligned} U\left(F(\hat{\mathbf{x}},\hat{\mathbf{s}})\right) &= \int\left(\int A\left(F(\cdot\,|\,\hat{\mathbf{s}})\right)F(\hat{\mathbf{s}}\,|\,\hat{\mathbf{x}})\,d\hat{\mathbf{s}}\right)p(\hat{\mathbf{x}})\,d\hat{\mathbf{x}} \\ &= \sum_{\forall\rho}\int v_{\rho}P(\rho\,|\,\hat{\mathbf{x}})\,p(\hat{\mathbf{x}})\,d\hat{\mathbf{x}} \end{aligned} \quad (16)$$

The transformation of Eq. (14) is achieved by considering the partial conditional probabilities that are the integral of the conditional probabilities $P(\rho\,|\,\hat{\mathbf{x}})$ over the variate vectors $\mathbf{x}^{\chi}$. By defining $\hat{\mathbf{x}}^{\tilde{\chi}} = \begin{bmatrix}\mathbf{x}^{1} & \cdots & \mathbf{x}^{\chi-1} & \mathbf{x}^{\chi+1} & \cdots\end{bmatrix}$, the partial conditional probabilities can be denoted as $P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)$, and this can be calculated as:

$$P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right) = \int P(\rho\,|\,\hat{\mathbf{x}})\,p(\mathbf{x}^{\chi})\,d\mathbf{x}^{\chi}. \quad (17)$$

Based on LEMMA 1, this study proves that the mutual information for the DM's prior and posterior knowledge for a variate vector $\mathbf{x}^{\chi}$ can be equivalently transformed to the mutual between $P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)$ and $P(\rho\,|\,\hat{\mathbf{x}})$ with the optimal information strategy. Eq. (14) can then be transferred as follows. See Appendix D for details.

$$\hat{c}(F) = \sum_{\forall\chi}\lambda^{\chi}\left(-\int H\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)p\left(\hat{\mathbf{x}}^{\tilde{\chi}}\right)d\hat{\mathbf{x}}^{\tilde{\chi}} + \int H\left(P(\rho\,|\,\hat{\mathbf{x}})\right)p(\hat{\mathbf{x}})\,d\hat{\mathbf{x}}\right). \quad (18)$$

3.3.2 Choice probability formulation

The MRI version of Eq. (8) can then be formed. Because both the ex ante expected utility and information cost can be denoted by the choice probabilities, only the choice probabilities $P(\rho\,|\,\hat{\mathbf{x}}), \forall\rho$ are taken as independent variables. See Appendix E for the proof. The formulation of $P(\rho\,|\,\hat{\mathbf{x}})$ can then be derived with the following optimization model.

$$\max_{P(\rho|\hat{\mathbf{x}})}\left(U\left(F(\hat{\mathbf{x}},\hat{\mathbf{s}})\right) - \hat{c}\left(F(\hat{\mathbf{x}},\hat{\mathbf{s}})\right)\right) \quad (19)$$

Subject to

$$\forall\rho:\ P(\rho\,|\,\hat{\mathbf{x}}) \geq 0 \quad (20)$$

$$\sum_{\forall\rho} P(\rho \mid \hat{\mathbf{x}}) = 1. \tag{21}$$

By using the Lagrange multiplier method presented in Appendix E to solve the above problem and letting $\lambda' = \sum_{\forall\chi} \lambda^{\chi}$, the formulation of the conditional choice probability is as follows:

$$P(\rho \mid \hat{\mathbf{x}}) = \frac{\exp(v_{\rho}/\lambda') \cdot \prod_{\forall\chi} P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}})^{\lambda^{\chi}/\lambda'}}{\sum_{\forall\rho'} \left(\exp(v_{\rho'}/\lambda') \cdot \prod_{\forall\chi} P(\rho' \mid \hat{\mathbf{x}}^{\tilde{\chi}})^{\lambda^{\chi}/\lambda'} \right)}. \tag{22}$$

Based on Eq. (22), the following properties of the MRI discrete choice model can be derived:

i. When the partial-conditional probabilities $P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}})$ are uniform, Eq. (22) reduces to the multinomial logit model.

ii. The bigger the value of λ' is, the less influence the variates have on the DM's decision.

iii. The bigger the value of λ^{χ} is, the more influence the partial conditional probabilities $P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}})$ have and the less influence the χ-th variate has on the DM's decision.

iv. Eq. (22) can be used to derive the choice probability formulation of the URI discrete choice model, i.e. Eq. (11), because partial conditional probabilities $P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}})$ are equivalent to unconditional probabilities $P(\rho)$ when the utilities of alternatives only depend on one variate.

3.3.3 *Solving the partial-conditional probability*

Eq. (22) does not explicitly express the choice probabilities because it depends on $P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}})$. By substituting Eq. (22) into the objective function Eq. (19), the MRI version of LEMMA 2 can be obtained, which can be used to solve for $P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}})$.

LEMMA 2: When $\forall\rho : P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}) > 0$ and the strategy $F(\hat{\mathbf{X}}, \hat{\mathbf{S}})$ is optimal, Eq. (23) holds.

$$\int \frac{\exp\left(v_{\rho}/\lambda'\right)\cdot \prod_{\forall\chi'\neq\chi} P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}'}\right)^{\lambda^{\chi'}/\lambda'} \cdot P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'-1}}{\sum_{\forall\rho'}\left(\exp\left(v_{\rho'}/\lambda'\right)\cdot \prod_{\forall\chi} P\left(\rho'\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)}\, p(\mathbf{x}^{\chi})d\mathbf{x}^{\chi} = 1. \tag{23}$$

For the proof, see Appendix F.

Eq. (23) is easy to understand by multiplying $P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)$ by both sides of the expression. The result ensures that the conditional probability integrates with the partial-conditional probability, as shown in Eq. (17). One can begin with an initial guess of these probabilities and iteratively update them to acquire the vector of partial-conditional probabilities $\tilde{\mathbf{P}}_{\rho} = \left[P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{1}}\right) \;\cdots\; P_{n,\rho}^{\sim 2}\left(\hat{\mathbf{x}}^{\tilde{\chi}}\right) \;\cdots\right]$ according to:

$$\left(\tilde{\mathbf{P}}_{\rho}\right)_{new} = L\left(\left(\tilde{\mathbf{P}}_{\rho}\right)_{old}\right) \odot \left(\tilde{\mathbf{P}}_{\rho}\right)_{old}, \tag{24}$$

where $L(\cdot)$ is defined as:

$$\begin{aligned} L\left(\tilde{\mathbf{P}}_{\rho}\right) &= \left[l_1\left(\tilde{\mathbf{P}}_{\rho}\right) \;\cdots\; l_{\chi}\left(\tilde{\mathbf{P}}_{\rho}\right) \;\cdots\right] \\ l_{\chi}\left(\tilde{\mathbf{P}}_{\rho}\right) &= \int \frac{\exp\left(v_{\rho}/\lambda'\right)\cdot \prod_{\forall\chi'\neq\chi} P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}'}\right)^{\lambda^{\chi'}/\lambda'} \cdot P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'-1}}{\sum_{\forall\rho'}\left(\exp\left(v_{\rho'}/\lambda'\right)\cdot \prod_{\forall\chi} P\left(\rho'\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)}\, p(\mathbf{x}^{\chi})d\mathbf{x}^{\chi} \end{aligned}. \tag{25}$$

According to the Blahut–Arimoto algorithm (Cover, 1999), the iteration based on Eq. (24) converges when the condition in Eq. (23) holds for all $\mathbf{x}^{\chi}$. In each iteration, the terms in $L\left(\left(\tilde{\mathbf{P}}_{\rho}\right)_{old}\right)$ indicate whether the terms in $\tilde{\mathbf{P}}_{\rho}$ need to be raised or dropped.

3.4 Illustrating the properties of the MRI discrete choice model

This section presents the numerical result for the example in Subsection 3.1 to illustrate the properties of the MRI discrete choice model. According to Eq. (22), the choice probabilities for the DM depend on the unit information costs λ^1, λ^2 and the utility coefficients β^1, β^2. To illustrate the impact of λ^1, λ^2, it is assumed that β^1, β^2 are given as $\beta^1 = -0.1, \beta^2 = -0.1$.

This study uses the format illustrated in Fig. 2 to present the value of conditional probability, particle-conditional probability, and state of variate vectors for a certain unit information value. Fig. 2 presents a dual-modal representation. The bar chart displays the particle-conditional probability $P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{1}}\right)=P\left(\rho \mid \mathbf{x}^{2}\right)$ and the pie charts illustrate the conditional probability $P(\rho \mid \hat{\mathbf{x}})$ that shares the axis that represents the state of $\mathbf{x}^2$. Similarly, the bar charts for the particle-conditional probability $P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{2}}\right)=P\left(\rho \mid \mathbf{x}^{1}\right)$ and the pie charts share the axis that represents the state of $\mathbf{x}^1$. The serial numbers in the pie chart footnotes aid in identification of pie charts.

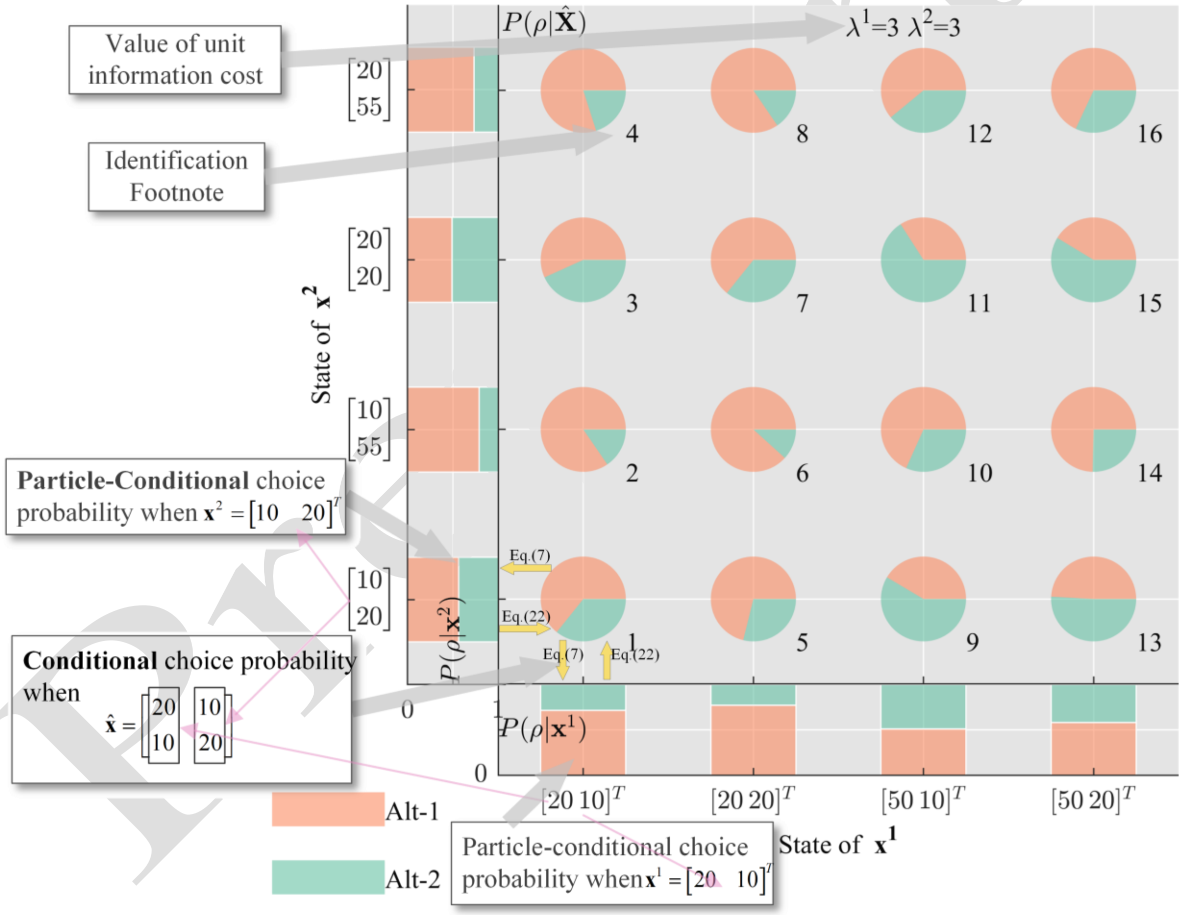


Fig. 2 Chart format showing conditional and particle-conditional probability distributions

Error! Reference source not found. depicts the choice probability distributions for different information costs, based on which the key properties of the MRI discrete choice model are determined.

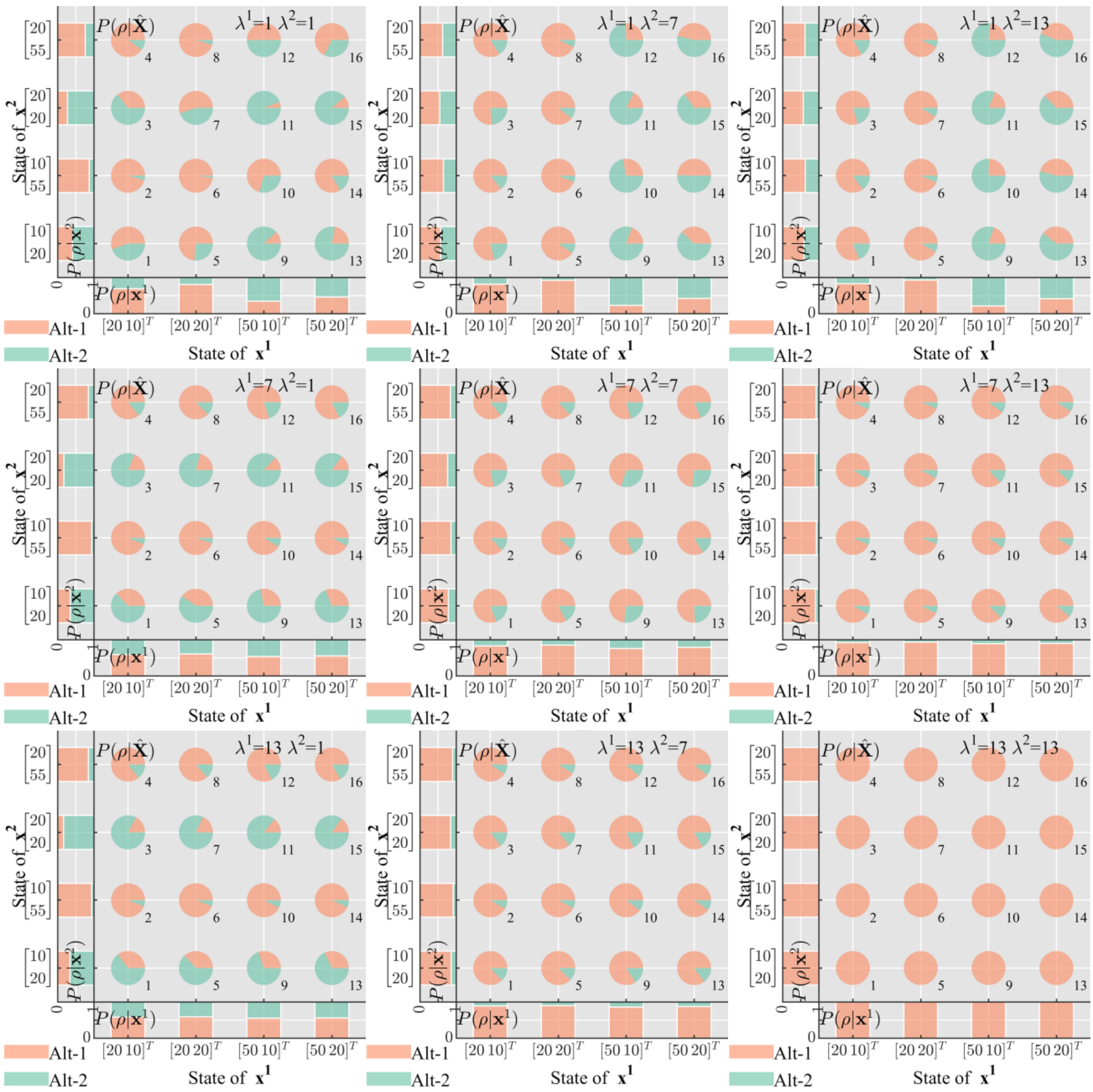


Fig. 3 Probability distributions for alternatives with different information costs

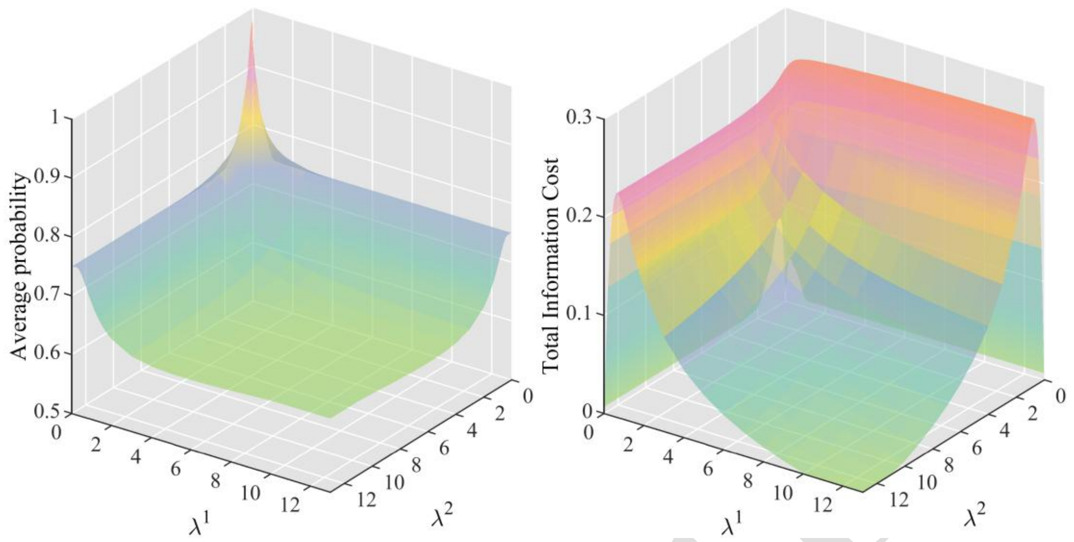
i. The states of the variates influence the choice probabilities. In the scenario depicted, for example, a unit information cost below the threshold of 13 means that any reduction in the state value of a given variate for any alternative heightens the related choice probability. This effect arises because the benefit of information in decision quality improvement can cover the information cost.

ii. As the unit information cost escalates, the influence of the true states of the attributes on the choice probabilities diminishes. To illustrate this, consider the pie charts labeled 6 and 14, for which the values of x_1^1 stand at 20 and 50, respectively. While the discrepancy in the choice probability for selecting an alternative between these charts is substantial at 0.4458, it plummets to a marginal 0.0498 when λ^1, λ^2 shifts from 1 and 7 to 7 and 7. The rationale for this is straightforward. The higher the unit information cost, the scantier the information the DM acquires, which diminishes the DM's responsiveness to the actual state of the attribute.

iii. The prior knowledge influences the choice probabilities. Taking pie chart labeled 1 as an illustration, two alternatives offer identical utility to the DM. However, the choice probabilities for the two alternatives differ significantly when $\lambda^1 = 1, \lambda^2 = 13$. This discrepancy stems from the DM's certainty that alternative 1 is at least as good as alternative 2, given the following: 1) The DM is aware of the inferiority of x_ρ^1 for alternative 1 relative to alternative 2, due to the smaller value of λ^1, 2) The state of $\mathbf{x}^2$ is unknown to the DM, which is attributed to the larger value of λ^2, and 3) The DM possesses prior knowledge that there is a potential for the x_ρ^2 of alternative 2 to be significantly inferior to that of alternative 1. A similar result can be found when $\lambda^1 = 13, \lambda^2 = 1$, which shows a contrary numerical result.

iv. As the unit information cost escalates, so does the influence of the prior knowledge on decision-making, and the probability of choosing an alternative with more potential to yield a higher utility increases. For example, the prospect that alternative 1 delivers a higher utility supersedes that of alternative 2. Subsequently, the probability of the DM selecting alternative 1 rises when λ^1, λ^2 shifts from 1 and 1 to 13 and 13, indicating a preference for options with a perceived higher benefit potential amidst rising information costs.

Fig. 4 depicts the average probability of the DM choosing the optimal alternative (a) and the average total information cost (b) aggregated from all potential states across varying unit information costs. An analysis of Fig. 4 reveals several key insights, which are listed below.



(a) (b)

Fig. 4 Average probability of choosing the best alternative (first) and total information cost (second) for different unit information costs

i. There is an inverse relationship between the unit cost of information and the propensity to choose the optimal alternative. As depicted in Fig. 4 (a), the average probability of choosing the best alternative approach, 1, when the value of unit information cost is small (such as $\lambda^1 \to 0, \lambda^2 \to 0$), is much higher than that when the information is expensive $\left(\lambda^1 = 13, \lambda^2 = 13\right)$.

ii. A discernible decline in the average probability of choosing the best option occurs abruptly as λ^1, λ^2 increases, and the probability thereafter moderates its descent when $\lambda^1 > 2, \lambda^2 > 2$. This phenomenon suggests that the marginal impact of the unit information cost wanes as information becomes more expensive. The primary reason for this is that the changes in the mutual information are not sensitive to the changes in the unit information cost when the information cost is expensive. Moreover, the information cost reaches its maximum at $\lambda^1 = 2, \lambda^2 = 2$ and then decreases, as shown in Fig. 4(b). This finding aligns with previous URI discrete choice model analyses conducted by Jiang et al. (2020) and Fosgerau and Jiang (2019). The cause for this may be that mutual information and unit information costs are inversely proportional if the total information cost is constant.

4 Application and Calibration

In this section, we explore the application of the MRI discrete choice model to characterize the route choice preferences of metro passengers. We begin by specifying the utility function, the prior knowledge, and the unit information cost, followed by a description of the devising of the calibration method. To facilitate clarity, Appendix G recapitulates the notations utilized herein, excepting the notations adapted and extended from Section 3. For example, to distinguish the choice situation among different passengers, the subscript n is added to the alternative route notation ρ.

4.1 Metro route choice

4.1.1 Utility function

When a passenger-n is planning to travel from the origin station (O) to the destination station (D) via a metro network, he/she can chooses one route from a set of available routes. When they travel via route ρ, the passenger may use a set of metro lines $\mathrm{L}_{n\rho} = \{l_{n\rho\tau} \mid \tau = 1, 2, \ldots\}$, and traverse a set of metro sections $\mathrm{K}_{n\rho} = \{k_{n\rho\kappa} \mid \kappa = 1, 2, \ldots\}$, which differentiates the available routes. Following Hörcher and Tirachini (2021), the utility associated with route ρ is defined by the following variates:

i. Number of transfers. For passenger-n traveling via route ρ, the number of transfers is denoted as $x_{n\rho}^{\text{trans}}$.

ii. Waiting time. The waiting time for passenger-n to board line $l_{n\rho\tau}$ is denoted by $x_{n\rho\tau}^{\text{wait}}$, and this waiting time is influenced by headways and delayed-boarding probability (Yap and Cats, 2021). Therefore, the total waiting time across all lines in route ρ is given by $x_{n\rho}^{\text{wait}} = \sum_{\tau=1}^{|\mathrm{L}_{n\rho}|} x_{n\rho\tau}^{\text{wait}}$.

iii. Walking time. The walking time before passenger-n boards line $l_{n\rho\tau}$ is denoted by $x_{n\rho\tau}^{\text{walk}}$, and this walking time is influenced by the distance between platforms or gate machines. Therefore, the total walking time across all lines in route in route ρ is given by $x_{n\rho}^{\text{walk}} = \sum_{\tau=1}^{|\mathrm{L}_{n\rho}|} x_{n\rho\tau}^{\text{walk}}$.

iv. Travel time. The travel time for passenger-n in section $k_{n\rho\kappa}$ is denoted by $x_{n\rho\kappa}^{\text{travel}}$, and this travel time is determined by the train operation schedule. Therefore, the total travel time across all sections in route ρ is given by $x_{n\rho}^{\text{travel}} = \sum_{\kappa=1}^{|\mathrm{K}_{n\rho}|} x_{n\rho\kappa}^{\text{travel}}$.

v. Expected standing time. The expected standing time for passenger-n on section $k_{n\rho\kappa}$ is denoted by $x_{n\rho\kappa}^{\text{stand}}$, as they may not occupy a seat in section $k_{n\rho\kappa}$. Therefore, the total expected standing time across all sections in route ρ is given by $x_{n\rho}^{\text{stand}} = \sum_{\kappa=1}^{|\mathrm{K}_{n\rho}|} x_{n\rho\kappa}^{\text{stand}}$.

vi. Crowding level. The crowding level for line section $k_{n\rho\kappa}$ is represented by the product of the standing passenger density $D_{n\rho\kappa}$ and the expected standing time $x_{n\rho\kappa}^{\text{stand}}$ (Hörcher et al., 2017). Therefore, the crowding level on route ρ is given by $x_{n\rho}^{\text{crowd}} = \sum_{\kappa=1}^{|\mathrm{K}_{n\rho}|} D_{n\rho\kappa} x_{n\rho\kappa}^{\text{stand}}$.

By using $x_{n\rho}^{\text{trans}}$, $x_{n\rho}^{\text{wait}}$, $x_{n\rho}^{\text{walk}}$, $x_{n\rho}^{\text{travel}}$, $x_{n\rho}^{\text{stand}}$, and to form the variate vectors $\mathbf{x}_n^{\chi}$ and variate matrix $\hat{\mathbf{x}}_n$, and using $\left[\beta^{\text{trans}} \quad \beta^{\text{wait}} \quad \beta^{\text{walk}} \quad \beta^{\text{travel}} \quad \beta^{\text{stand}} \quad \beta^{\text{crowd}}\right]$ to form the linear utility payoff coefficient vector $\boldsymbol{\beta}$, the utility vector $\mathbf{v}_n$ corresponding to the route set of passenger-n can be denoted with Eq. (26).

$$\begin{aligned}\mathbf{v}_n &= \hat{\mathbf{x}}_n \boldsymbol{\beta}^T \\ &= \left[\mathbf{x}_n^{\text{trans}} \quad \mathbf{x}_n^{\text{wait}} \quad \mathbf{x}_n^{\text{walk}} \quad \mathbf{x}_n^{\text{travel}} \quad \mathbf{x}_n^{\text{stand}} \quad \mathbf{x}_n^{\text{crowd}}\right] \cdot \left[\beta^{\text{trans}} \quad \beta^{\text{wait}} \quad \beta^{\text{walk}} \quad \beta^{\text{travel}} \quad \beta^{\text{stand}} \quad \beta^{\text{crowd}}\right]^T\end{aligned} \tag{26}$$

4.1.2 Prior knowledge

Implementing the MRI discrete choice model necessitates defining the DM's prior knowledge. In the context of the metro route choice problem, the prior knowledge of passengers can be inferred from the raw distributions that reflect real-world conditions. This premise follows an approach previously validated by Jiang et al. (2020). Furthermore, this study assumes that the variates mentioned in Section 4.1 are independent of each other. This assumption is consistent with those made in MRI models applied to other problems. (e.g., Miao et al. (2022), Peng and Xiong (2006), Van Nieuwerburgh and Veldkamp (2010), and Zorn (2020)). In the following discussion, we elaborate on how the raw distributions for different variates are obtained.

The raw distributions for the number of transfers, walking time, and travel time represent the variation in each of these attributes across different OD pairs. The raw distribution captures two key components: the state space (all possible variate combinations) and the probability assigned to each state. The state space is constructed by enumerating all possible combinations of the variate levels across the various routes for each OD pair. For the probability, each combination is assigned a probability based on its frequency of occurrence across all possible routes.

The raw distributions of waiting time, standing time, and crowding levels capture the variations in each of these factors throughout the daily operation of the metro system. These distributions are derived by evaluating the states of the route set between any given origin-destination (OD) pair through the following procedure: 1) Compilation of State Combinations: For each variate, the state combinations, representing the varying states across different routes in the route set, are collected at different times during operational hours. 2) Discretization of State Space: The state space for each variate is discretized into uniform intervals. The intervals corresponding to the states at various times are then identified for each variate. 3) Median State Assignment: The actual state of each route for each variate is replaced by the median state within each identified interval. 4) Probability Assignment: A probability is assigned to each combination of discrete states for each variate, proportional to its frequency of occurrence.

4.1.3 *Unit information cost*

Before reaching a decision, passengers have different ways to acquire information about variates, including network schematic diagrams, passenger information systems, social media, and real-time information apps. The difficulty of seeking and understanding information with different variates varies. For example, the number of transfers is easily identifiable and understandable from network schematic diagrams, whereas crowding levels are typically retrieved from real-time information apps.

Additionally, the challenge of seeking and processing information for a specific variate may differ among different OD pairs. In attempting to seek information about walking, waiting, travel, standing time, and crowding level, a passenger must process this information in an integrated manner for each transit line or section. Consequently, the greater the number of lines and sections encompassing an OD pair, the more effort is required for a passenger to acquire the information. Hence, it is proposed that the unit information costs associated with these factors are proportional to the number of transit lines or segments, as follows:

$$\begin{cases} \lambda_n^{\text{walk}} = \sum_{\forall \rho} \left| \mathrm{L}_{n\rho} \right| \lambda^{\text{walk}} \\ \lambda_n^{\text{wait}} = \sum_{\forall \rho} \left| \mathrm{L}_{n\rho} \right| \lambda^{\text{wait}} \\ \lambda_n^{\text{travel}} = \sum_{\forall \rho} \left| \mathrm{K}_{n\rho} \right| \lambda^{\text{travel}} \\ \lambda_n^{\text{stand}} = \sum_{\forall \rho} \left| \mathrm{K}_{n\rho} \right| \lambda^{\text{stand}} \\ \lambda_n^{\text{crowd}} = \sum_{\forall \rho} \left| \mathrm{K}_{n\rho} \right| \lambda^{\text{crowd}} \end{cases}, \tag{27}$$

where λ_n^{walk}, λ_n^{wait}, $\lambda_n^{\text{travel}}$, λ_n^{stand}, and λ_n^{crowd} represent the unit information cost for the corresponding variates of passenger-*n*, and λ^{walk}, λ^{wait}, λ^{travel}, λ^{stand}, and λ^{crowd} represent the basic unit information cost associated with corresponding variates when $\sum_{\forall \rho} \left| \mathrm{L}_{n\rho} \right| = 1$ and $\sum_{\forall \rho} \left| \mathrm{K}_{n\rho} \right| = 1$.

4.2 Model calibration

For the observed choice $y_{n\rho}$, the states of variates $\hat{\mathbf{x}}_n$ and the distribution of the prior knowledge distribution $p(\hat{\mathbf{x}}_n)$ can be ascertained from a sample of *N* passengers for the purpose of model calibration. Then a maximum likelihood estimation technique can be employed to estimate the utility coefficient vector $\boldsymbol{\beta}$ and the basic unit information cost vector $\boldsymbol{\lambda} = \left[\lambda^{\text{trans}}, \lambda^{\text{walk}}, \lambda^{\text{wait}}, \lambda^{\text{travel}}, \lambda^{\text{stand}}, \lambda^{\text{crowd}} \right]$.

The probability of passenger-*n* choosing the route they have observed can be expressed as $\prod_{\forall \rho} \left(P(\rho \mid \hat{\mathbf{x}}_n) \right)^{y_{n\rho}}$, where $y_{n\rho} = 1$ denotes passenger-*n*'s chosen route ρ, and otherwise $y_{n\rho} = 0$. With the assumption that the route choice of each passenger is independent, the probability of every individual in the sample opting for their observed route is computed as:

$$\prod_{n=1}^{N} \prod_{\forall \rho} \left(P(\rho \mid \hat{\mathbf{x}}_n) \right)^{y_{n\rho}} . \tag{28}$$

Then the log-likelihood function of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ is specified by:

$$LL(\boldsymbol{\beta},\boldsymbol{\lambda})=\sum_{n=1}^{N}\sum_{\forall\rho}y_{n,\rho}\ln\left(P\left(\rho\,|\,\hat{\mathbf{x}}_n\right)\right). \tag{29}$$

The unknown parameters in Eq. (29) are the utility coefficient vector $\boldsymbol{\beta}$, the base unit information cost vector $\boldsymbol{\lambda}$, and the partial conditional probabilities $P\left(\rho\,|\,\hat{\mathbf{x}}_n^{\tilde{z}}\right)$. According to subsection 3.3.3, the partial conditional probabilities $P\left(\rho\,|\,\hat{\mathbf{x}}_n^{\tilde{z}}\right)$ can be determined by iteration using Eq. (25) according to $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and $p\left(\hat{\mathbf{x}}_n\right)$. Thus, the calibration problem is a bilevel optimization problem. The goal for the upper level is to maximize Eq. (29) by optimizing $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, while the lower level is utilized to compute the partial conditional probabilities $p\left(\hat{\mathbf{x}}_n\right)$ corresponding to the $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$ given from the upper level. This optimization problem can be solved using the Optimization Toolbox in MATLAB.

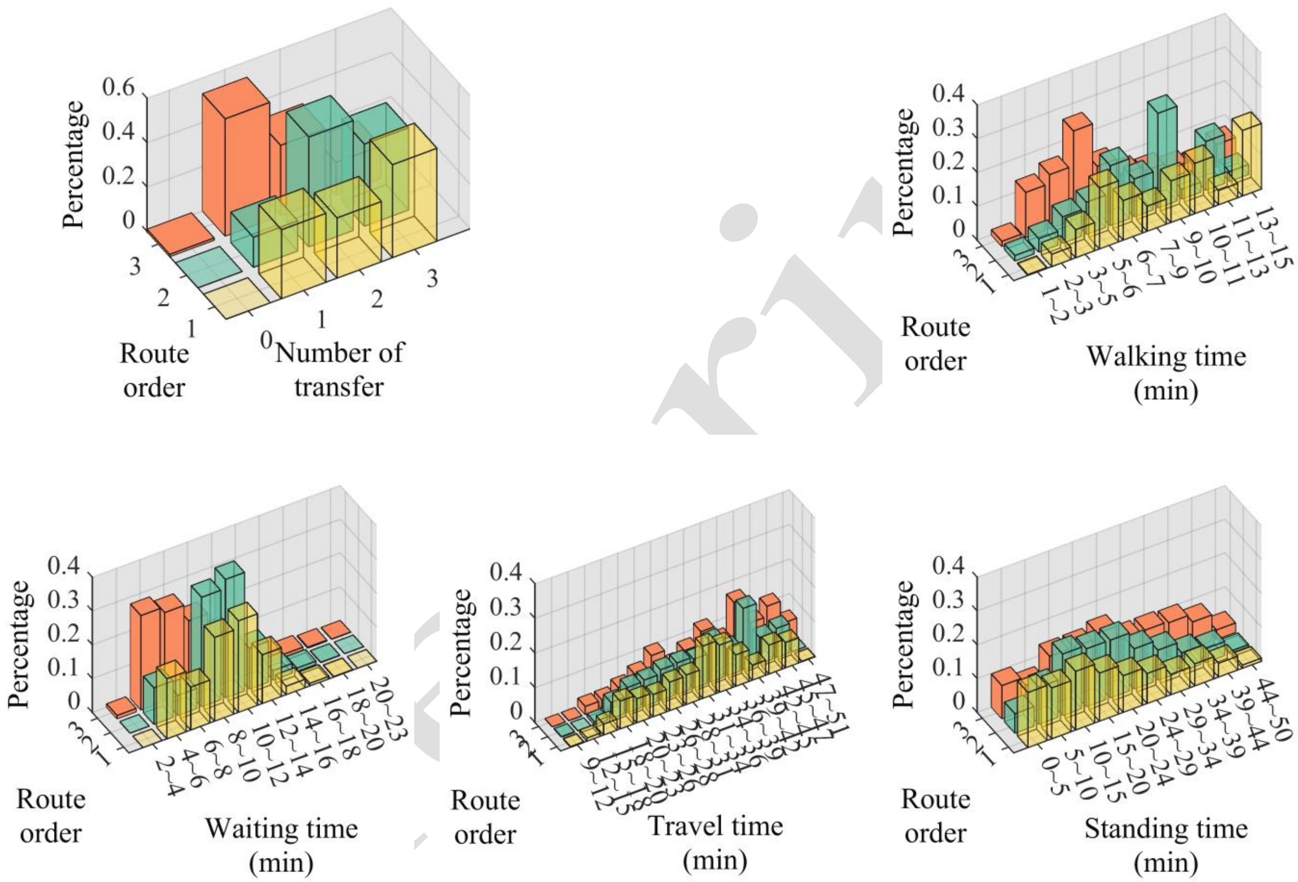
5 Case Study

5.1 Data Description

The case study is set against the backdrop of the Chengdu metro. At the time the case data were collected, the Chengdu metro comprised six lines and 136 stations, offering passengers multiple route alternatives. Passengers can access information about the routes through various means, including real-time map apps (e.g., Amap, Baidu Map), passenger information boards, and network schematic diagrams.

The observed choices of passengers, the states of the route set, and the raw distribution for each variate were collected from Chengdu Metro by Chen et al. (2023). This collection was based on an itinerary estimation method that inferred passenger space-time trajectories using smart card data. The data set contains 980,787 alternatives for 370,937 passengers.

The sample exhibits an uneven distribution across different origin-destination (OD) pairs, which may introduce potential biases into the analysis. Additionally, the large dataset of 370,937 records would compromise computational efficiency. To address these issues, we select a subset of records by prioritizing orthogonality and balance in the attribute levels. The data screening was conducted for origin-destination (OD) pairs with three alternative routes, because most OD pairs in the dataset have two or three available routes. Additionally, the study by Rolfe and Bennett (2009) suggests that a three-alternative choice problem provides a more robust model. The goal of this approach is to enhance the representativeness and analytical manageability of the sample while mitigating computational constraints. The screened sample includes 22,342 passengers who travel between 455 OD pairs.



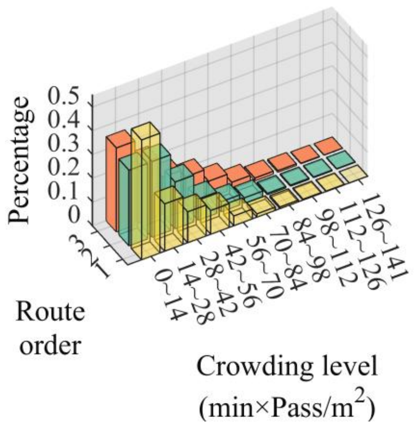


Fig. 5 presents a series of six bar plots that illustrate the distribution of various route variates across three different routes. In each plot, the yellow, green and orange bars represent the distribution associated with the first, second, and third route, respectively. The three routes are ordered by length, facilitating a clear comparison of the distribution of each variate across the different routes.

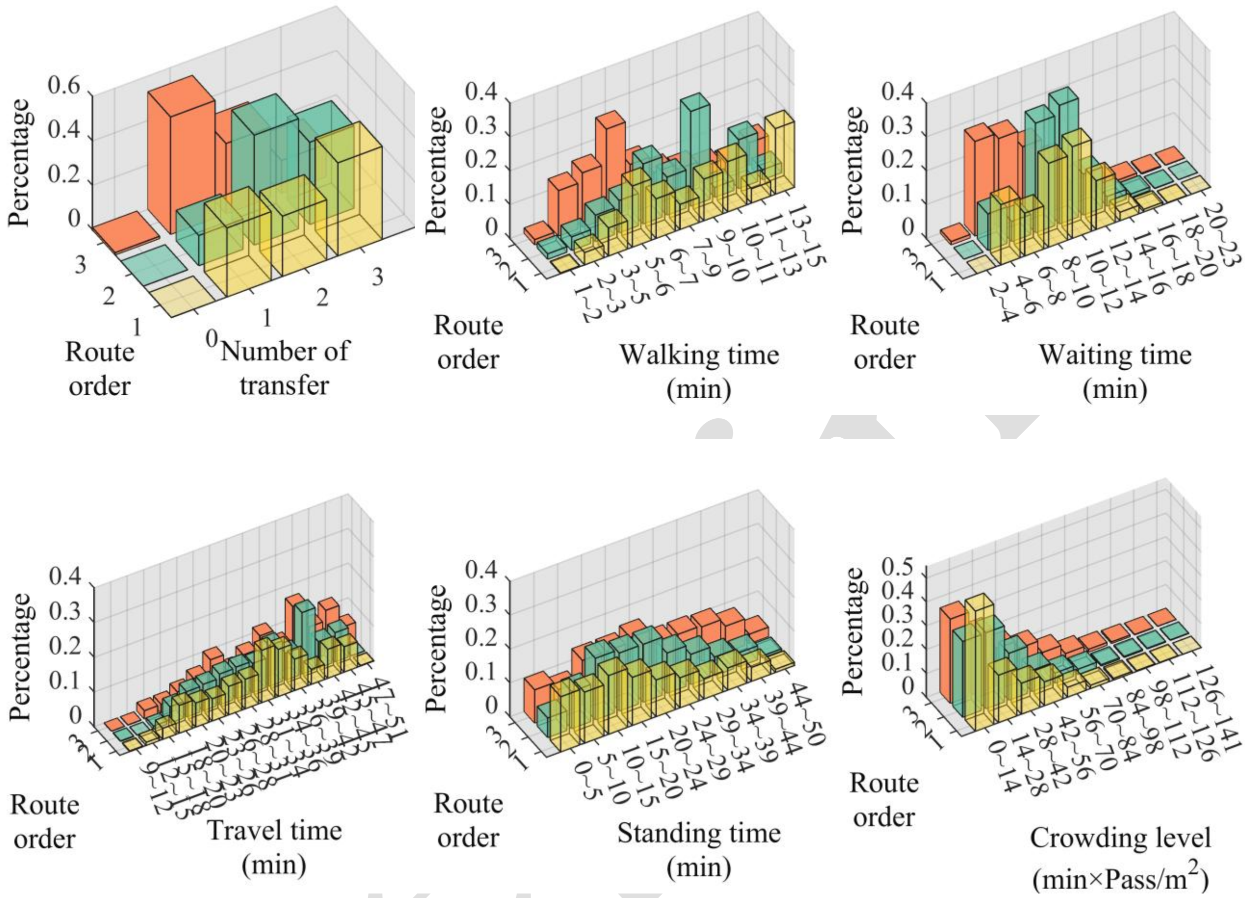


Fig. 5 Distribution of variate levels for different routes

Following the procedure outlined in sub-section 4.1.2, the raw distributions for various route variates were calculated. For the number of transfers, walking time, travel time, there are 24, 56, 123 distinct combinations. For waiting time, standing time and crowding level, the length of uniform intervals for each variate are 0.5 minutes, 4 minutes, and 4 (min × Pass/m²), and the maximum number of combinations among the sample OD pairs are 85, 129 and 151.

5.2 Calibration results

This section presents the calibration results from the MRI discrete choice model and the benchmarking for the results based on the results obtained from the URI, MNL and ANA models. The MNL model is selected because it is currently the mainstream model for studying the route choice preference of metro passengers (Hörcher et al., 2017, Li and Hensher, 2011, Yap et al., 2018), and this model assumes that the DMs are fully informed. In addition, based on the calibration results shown in Table 3, we also calculate the willingness to pay (WTP) associated with different travel time components. This is presented in Table 4.

Table 3 Calibration results for the MRI, URI, MNL, and ANA discrete choice model

MRI				URI				MNL		ANA			
β^{trans}	−0.4035 *** (0.0026)	λ^{trans}	0.0073 *** (0.0003)	β^{trans}	−7.9768 (0.7513)	λ	3.7913 (0.7663)	β^{trans}	−0.7466 *** (0.0090)	β^{trans}	−1.9086 *** (0.0007)	λ^{trans}	14.6299 (0.7007)
β^{walk}	−0.3528 *** (0.0072)	λ^{walk}	2.1751 *** (0.0164)	β^{walk}	−0.0417 (0.2551)			β^{walk}	−0.4216 *** (0.0087)	β^{walk}	−3.4315 *** (0.0001)	λ^{walk}	0.0043 *** (0.0036)
β^{wait}	−0.0145 (0.0251)	λ^{wait}	46.1543 (15.4218)	β^{wait}	−10.2971 (0.5060)			β^{wait}	−0.0158 (0.0289)	β^{wait}	0.3649 *** (0.0086)	λ^{wait}	0.3877 *** (0.0098)
β^{travel}	−0.0754 *** (0.0091)	λ^{travel}	0.0017 *** (0.0001)	β^{travel}	−2.8683 (0.6991)			β^{travel}	−0.2464 *** (0.0057)	β^{travel}	−0.2422 *** (0.0029)	λ^{travel}	19.6577 *** (0.0003)
β^{stand}	−0.0144 *** (0.0008)	λ^{stand}	0.2265 *** (0.0126)	β^{stand}	1.3225 (0.8909)			β^{stand}	−0.0136 *** (0.0028)	β^{stand}	−0.5004 *** (0.0076)	λ^{stand}	0.0000 (0.7017)
β^{crowd}	−0.0043 *** (0.0002)	λ^{crowd}	0.3324 *** (0.0144)	β^{crowd}	9.3183 (0.9593)			β^{crowd}	0.0174 *** (0.0016)	β^{crowd}	0.0220 *** (0.0062)	λ^{crowd}	0.4115 (0.1219)
LL	−13,535.9432			−91,564.2164				−13,721.8724		−14398.5027			
R^2	0.4434			−2.7304				0.4358		0.4134			

Note: Std. errors in brackets $*p < 0.1$; $** p < 0.05$; $*** p < 0.01$; LL for log-likelihood; R^2 for McFadden's pseudo-R^2

Table 4 Willingness to pay (WTP)

	Utility for MRI (minute/minute)	Unit information cost for MRI (bit/minute)	Utility for MNL (minute/minute)
Number of transfers	5.3515	0.0968	3.0300
Walking time	4.6790	28.8470	1.7110
Travel time	1.0000	0.0225	1.0000
Standing time	0.1910	3.0040	0.0552
Crowding level	0.0570	4.4085	−0.0706

Note: WTP is computed as the ratio of travel time coefficient for MRI and MNL

5.2.1 *Calibration results for the MRI discrete choice model*

The results in Table 3 and Table 4 from the MRI discrete choice model substantiate the following expectations.

i. The negative signs of the estimated coefficients reveal that the number of transfers, walking time, waiting time, travel time, standing time, and crowding all contribute negatively to utility.

ii. The coefficient and the unit information cost for the waiting time are insignificant. The reason for this could be that short waiting times have a minor impact on utility (Nielsen et al., 2021).

iii. The WTP of the walking time is 4.6790, which means that one unit of walking time brings the same disutility as that of 4.6790 units of travel time. The result accords with the reality that passengers dislike spending time walking more than spending time traveling.

iv. The WTP of the standing time is 0.1910, indicating that one unit of standing time is equivalent to an additional disutility of 0.1910 units of travel time. This value is notably lower than the previously reported figures based on stated preference data, such as the 1.53 value measured by Whelan and Crockett (2009). This discrepancy may be attributed to the differences between stated choices and actual behavior, which could lead to an overestimation of coefficient values when using stated preference data (Yap et al., 2018).

v. The WTP of the crowding level is 0.0570, which means that an additional passenger per square meter on average adds the additional disutility of 0.0570 units of travel time. This value is also lower than the previously reported figures based on stated preference data, such as the 0.085 measured by Whelan and Crockett (2009). The reasons for this discrepancy are similar to those discussed regarding the standing time.

vi. The basic unit information cost varies significantly across different variates, emphasizing the need to distinguish between the difficulties of information acquisition among these variates. However, this measure alone cannot reliably compare the overall difficulty of information acquisition across variables. For example, passengers may need to acquire only a small amount of information for some variables due to the low entropy of their raw distribution, e.g., the number of transfers. Consequently, a variable that has a high unit information cost but requires minimal information could result in a lower total information acquisition cost for passengers.

5.2.2 Calibration results for the URI discrete choice model

To conduct a comparative analysis, we calibrate the URI discrete choice model. We treat the utility for each alternative as univariate, following the method in Habib (2023), but we use the iterative method of Caplin et al. (2019) to obtain the unconditional choice probability. Since the method in Habib (2023) relies on market share to represent the unconditional choice probability, it may not be applicable in some cases, such as transit assignment. In Table 3, β^{trans}, β^{wait}, β^{walk}, β^{travel}, β^{stand}, and β^{crowd} in the URI discrete choice model are utility coefficients associated with different variates, and λ represents the unit information cost for the utility.

The results show that the likelihood value of the URI discrete choice model is −91,564.2164, with all variates statistically insignificant, which is significantly worse than that for the MRI discrete choice model. This performance issue arises because the URI model does not account for differences in the information acquisition difficulty among variates, which inherently means that it also fails to account for differences in the information acquisition difficulty among different OD pairs. The unit information cost that is suitable for some OD pairs may be either too high or too low for other OD pairs, leading to unconditional and conditional choice probabilities similar to those observed in Jiang et al. (2020) (see Fig. 7 in their paper). This suggests that passengers may only choose the route with the highest potential to be optimal when the unit information cost is high or may only select the optimal route when the unit information cost is low. Intuitively, as the number of metro lines or sections affecting a variate increases, the difficulty for passengers to acquire information for that variate rises, resulting in a higher unit information cost.

5.2.3 *Calibration results for the MNL model*

Table 3 and Table 4 also list the results for the MNL model. Drawing on Hensher et al. (2005), a pseudo-R^2 of 0.3 is deemed to be indicative of a respectable fit within the context of discrete choice models. The values of R^2 for the MNL and MRI discrete choice models are similar and much bigger than 0.3, indicating a similar fitness level.

However, the coefficient of the crowding level for the MNL model is positive, which means that the worse the crowding level of a route is, the more the passengers prefer it. The result is biased from intuition and thus lacks interpretability. The reason for this is that the MNL model mentioned above assumes that the DMs are fully informed of each variate (Hörcher et al., 2017, Li and Hensher, 2011, Wardman and Whelan, 2011, Yap and Cats, 2021, Yap et al., 2018).

Table 5 An example alternative route set

$y_{n,\rho}$	$r_{n,\rho}$	$x_{n,\rho}^{\text{trans}}$	$x_{n,\rho}^{\text{walk}}$	$x_{n,\rho}^{\text{wait}}$	$x_{n,\rho}^{\text{travel}}$	$x_{n,\rho}^{\text{stand}}$	$x_{n,\rho}^{\text{crowd}}$

1	1	1	3.9101	5.6467	42.6998	29.3999	39.5427
0	2	3	9.2849	11.2200	43.5998	16.6045	14.8116
0	3	2	7.2782	8.5537	45.8498	11.2888	3.4219

We then demonstrate how the fully informed assumption leads to biased coefficient results. Table 5 shows passenger choices and the states of variates in a typical scenario from the Chengdu Metro dataset. The table reveals that passengers prefer the shortest route, despite its long standing duration and unfavorable congestion levels, to alternatives. This type of scenario is not isolated. Similar scenarios account for 12.7520% of our dataset.

Observations from Table 5 indicate that increasing the utility coefficient for crowding raises the utility of the first route, thereby increasing the probability of choosing this route (as detailed in the study by Hensher et al. (2005)). Consequently, this adjustment tends to produce a larger crowding level coefficient in the likelihood estimation.

However, for passengers similar to those in Table 5, who may be unaware of crowding levels and standing time due to the difficulty of seeking information, the increase of the utility coefficient for crowding does not effectively explain passenger behavior. This highlights the fact that the fully informed assumption fails to account for the actual decision-making process of passengers, leading to biased results.

Analogous trends are observed for other variates as well, including the number of transfers, standing time, and crowding levels, revealing relative disparities in the utility WTP of 38.9642%, 69.5364%, and 240.3579%, respectively, between the MNL and MRI discrete choice models. Notably, a surge in the WTP associated with the basic unit of information amplifies these discrepancies because higher information costs tend to reduce passenger consideration for a given variate.

5.2.4 Calibration results for the ANA model

This section presents the calibration results for the ANA model. The ANA model used in this study follows(Hensher and Rose, 2009), which combines a discrete choice model with a non-attendance parameter to account for the possibility that some passengers may ignore certain attributes when making decisions. This model was applied to our RP dataset in contrast to the SP data used in their study. The model incorporates both the utility coefficients (β^{trans}, β^{wait}, β^{walk}, β^{travel}, β^{stand}, β^{crowd}) for each attribute and the non-attendance parameters (λ^{trans}, λ^{wait}, λ^{walk}, λ^{travel}, λ^{stand}, λ^{crowd}), which represent the probability that a DM does not consider a particular attribute. The following analysis discusses the calibration results and compares the ANA model to other models in the study.

i. The calibration results presented in Table 3 show that the ANA model has a likelihood value of −14,398.5027, which indicates a lower fit compared to the MNL model. This suggests that the ANA model is less effective in explaining the observed choices in our dataset. Compared to (Hensher and Rose, 2009), where the ANA model provided a better fit than the MNL model, our results indicate a notable difference in the model performance. The discrepancy may stem from the data type used in our study—RP data—while (Hensher and Rose, 2009) used SP data. Each passenger in the SP survey could contribute a series of observations. In contrast, in our dataset, each passenger contributes only one observation, which may limit the ability of the ANA model to capture the probability of non-attendance for each attribute accurately.

ii. The fit of the ANA model is lower than that of the MRI model. One possible reason is that the ANA model assumes that DMs either attend to or ignore specific attributes entirely, simplifying the decision-making process. However, in real-world scenarios, passengers' awareness of attributes often falls into an imperfect state, where they may have partial knowledge of an attribute, influenced by factors such as prior experience, imperfect information, or biases. Such a nuanced state of partial awareness is difficult for the ANA model to capture, as it does not account for intermediate levels of attribute attention. This limitation may reduce the model's ability to fully represent passenger behavior, particularly in situations where information is incomplete or imprecise. As pointed out by (Kravchenko, 2014), while the ANA model offers valuable insights, it may fall short in accurately modeling DMs' information seeking processes in more complex, real-world contexts.

iii. As observed in the ANA model, the positive utility coefficient for crowding implies that passengers may prefer routes with higher levels of crowding, which contradicts intuitive expectations. The reason for this is likely the same as in the MNL model: the increasing utility coefficient for crowding raises the utility of the first route, thereby inflating the probability of choosing that route.

6 Concluding Remarks

In this study, a model is devised to tackle the multivariate discrete choice with rational inattention, and this model is named the MRI discrete choice model. The MRI discrete choice model considers decision-makers to be uncertain about each alternative's state and to have access to the information to support their decision. The influence of information on choice preference is incorporated in the model by considering the information cost for each variate. The information cost is measured using the Shannon entropy-based information quantization and unit information cost. It is determined that the MRI discrete choice model results in probabilistic choices that follow a logit format and capture the influence of each variate's state and unit information cost corresponding to the prior knowledge of the DMs. At the same time, the unit information cost impacts the influence of the variate states and the prior knowledge of the DMs regarding choice probabilities. Specifically, the influence of a variate's state on the choice probabilities increases as the corresponding unit information cost decreases. In contrast, the influence of the DM's prior knowledge on the variate decreases with the corresponding unit information cost decreases.

Furthermore, in this study, the MRI discrete choice model is applied to the metro route choice problem, and the model is calibrated with the revealed route choice preference data collected from Chengdu Metro. To the best of the authors' knowledge, our study is the first to calibrate a discrete choice model with the rational inattention theory using real case data. The calibration results show that: 1) The willingness to pay for information varies significantly among different variates. 2) The number of transfers, walking time, waiting time, travel time, standing time, and crowding cause the disutility of passengers. 3) The influence of the utility coefficient and the unit information cost of waiting time is insignificant, which is in line with previous studies (e.g., Nielsen et al. (2021)).

By comparing the proposed model result with that of the URI discrete choice model, it is found that the resulting goodness of fit for the URI discrete choice model is unsatisfactory when market share is unavailable because the model cannot distinguish the information cost for different variates and OD pairs.

By comparing the results for the proposed model with those for the MNL model, it is found that: 1) The MRI discrete choice model and the MNL model perform similarly in terms of fitness for our data set. 2) The fully informed assumption in analyzing the revealed preference data via the MNL model may lead to an overestimation of the utility coefficients, which is avoided in the MRI discrete choice model.

By comparing the results for the proposed model with those for the ANA model, it is found that the fit of the ANA model is lower than that of the MRI model, and the ANA model may fall short in accurately modeling DMs' information seeking processes in real-world contexts.

There are a few limitations worth mentioning, and these are left for future research. First, similar to multivariate models with RI in other areas, such as Miao et al. (2022), Peng and Xiong (2006), Van Nieuwerburgh and Veldkamp (2010), Zorn (2020), we did not consider the variate dependence in the MRI discrete choice model. This type of simplification might lead to calibration bias, requiring future effort to resolve it. Second, the preference heterogeneity among the DMs is not considered in this study, but this could be settled by considering the parameters (that is, the utility coefficient and unit information cost) to be randomly distributed, similar to the mixed logit model. Third, due to the increased complexity of the proposed model, it requires significantly more computational time compared to the MNL model. For instance, calibration of the MRI dataset took 152,354 seconds for our data set, while MNL required only 378 seconds, highlighting the trade-off between interpretability and solution efficiency. Finally, this study only examines the model performance using metro data. The advantages of the model could be further explored using data from other transport systems such as buses, ride-sharing, or bicycles.

References

Amirgholy, M., Shahabi, M., Gao, H.O., 2017. Optimal design of sustainable transit systems in congested urban networks: A macroscopic approach. *Transportation Research Part E: Logistics and Transportation Review* 103, 261-285.

Ben-Elia, E., Avineri, E., 2015. Response to Travel Information: A Behavioural Review. *Transport Reviews* 35(3), 352-377.

Björklund, G., Swärdh, J.-E., 2017. Estimating policy values for in-vehicle comfort and crowding reduction in local public transport. *Transportation Research Part A: Policy and Practice* 106, 453-472.

Caplin, A., Dean, M., Leahy, J., 2019. Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies* 86(3), 1061-1094.

Chen, X., Jiang, Y., Bláfoss Ingvardson, J., Luo, X., Anker Nielsen, O., 2023. I can board, but I'd rather wait: Active boarding delay choice behaviour analysis using smart card data in metro systems. *Transportation Research Part A: Policy and Practice* 174.

Chorus, C.G., Molin, E.J., van Wee, B., 2006. Travel information as an instrument to change cardrivers' travel choices: a literature review. *European Journal of Transport and Infrastructure Research* 6(4).

Cover, T.M., 1999. *Elements of information theory*. John Wiley & Sons.

De Palma, A., Lindsey, R., Picard, N., 2012. Risk aversion, the value of information, and traffic equilibrium. *Transportation Science* 46(1), 1-26.

Ettema, D., Timmermans, H., 2006. Costs of travel time uncertainty and benefits of travel time information: Conceptual model and numerical examples. *Transportation Research Part C: Emerging Technologies* 14(5), 335-350.

Fosgerau, M., Jiang, G., 2019. Travel time variability and rational inattention. *Transportation Research Part B: Methodological* 120, 1-14.

Fosgerau, M., Melo, E., De Palma, A., Shum, M., 2020. Discrete choice and rational inattention: A general equivalence result. *International economic review* 61(4), 1569-1589.

Habib, K.N., 2023. Rational inattention in discrete choice models: Estimable specifications of RI-multinomial logit (RI-MNL) and RI-nested logit (RI-NL) models. *Transportation Research Part B: Methodological* 172, 53-70.

Hensher, D., 2014. Attribute processing as a behavioural strategy in choice making, in: Hess, S., Daly, A. (Eds.), *Handbook of Choice Modelling*. Edward Elgar Publishing, pp. 268-289.

Hensher, D.A., Greene, W.H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical economics* 39, 413-426.

Hensher, D.A., Rose, J.M., 2009. Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. *Transportation Research Part E: Logistics and Transportation Review* 45(4), 583-590.

Hensher, D.A., Rose, J.M., Greene, W.H., 2005. *Applied choice analysis: a primer*. Cambridge University Press.

Hogarth, R.M., Wiley, 1980. Judgment and choice: the psychology of decision. *Wiley*.

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

Hörcher, D., Tirachini, A., 2021. A review of public transport economics. *Economics of Transportation* 25, 100196.

Huang, D., Liu, Z., Liu, P., Chen, J., 2016. Optimal transit fare and service frequency of a nonlinear origin-destination based fare structure. *Transportation Research Part E: Logistics and Transportation Review* 96, 1-19.

Jiang, G., Fosgerau, M., Lo, H.K., 2020. Route choice, travel time variability, and rational inattention. *Transportation Research Part B: Methodological* 132, 188-207.

Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L., 2016. A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica* 84(2), 571-626.

Kravchenko, A.J.J.o.c.m., 2014. Influence of rudimentary attribute non-attendance (ANA) on choice experiment parameter estimates and design efficiency: A Monte Carlo Simulation analysis. 11, 57-68.

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

Maćkowiak, B., Matejka, F., Wiederholt, M., 2020. Rational Inattention: A Review. *CEPR Discussion Papers*(15408).

Matějka, F., McKay, A., 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review* 105(1), 272-298.

Miao, J., Wu, J., Young, E., 2019. Multivariate rational inattention. Boston University-Department of Economics.

Miao, J., Wu, J., Young, E.R., 2022. Multivariate rational inattention. *Econometrica* 90(2), 907-945.

Nielsen, O.A., Eltved, M., Anderson, M.K., Prato, C.G., 2021. Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice* 147, 76-92.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. *Journal of Financial Economics* 80(3), 563-602.

Rolfe, J., Bennett, J., 2009. The impact of offering two versus three alternatives in choice modelling experiments. 68(4), 1140-1148.

Sims, C.A., 2003. Implications of rational inattention. *Journal of Monetary Economics* 50(3), 665-690.

Sims, C.A., 2010. Rational Inattention and Monetary Economics. Elsevier, pp. 155-181.

Tirachini, A., Hensher, D.A., Jara-Díaz, S.R., 2010. Restating modal investment priority with an improved model for public transport analysis. *Transportation Research Part E: Logistics and Transportation Review* 46(6), 1148-1168.

Van Nieuwerburgh, S., Veldkamp, L., 2010. Information acquisition and under-diversification. *The Review of Economic Studies* 77(2), 779-805.

Wardman, M., Whelan, G., 2011. Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience. *Transport Reviews* 31(3), 379-398.

Whelan, G., Crockett, J., 2009. An investigation of the willingness to pay to reduce rail overcrowding, *Proceedings of the first International Conference on Choice Modelling, Harrogate, England*. Citeseer.

Yap, M., Cats, O., 2021. Taking the path less travelled: Valuation of denied boarding in crowded public transport systems. *Transportation Research Part A: Policy and Practice* 147, 1-13.

Yap, M., Cats, O., Van Arem, B., 2018. Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science* 16(1), 23-42.

Yin, H., Wu, J., Liu, Z., Yang, X., Qu, Y., Sun, H., 2019. Optimizing the release of passenger flow guidance information in urban rail transit network via agent-based simulation. *Applied Mathematical Modelling* 72, 337-355.

Zhu, Z., Li, X., Liu, W., Yang, H., 2019. Day-to-day evolution of departure time choice in stochastic capacity bottleneck models with bounded rationality and various information perceptions. *Transportation Research Part E: Logistics and Transportation Review* 131, 168-192.

Zorn, P., 2020. Investment under rational inattention: Evidence from us sectoral data.

Appendix

Appendix A: Notations in Section 2–3

ρ	Alternative of the DM, $\rho = 1, 2, \dots$
v_ρ	Utility payoff of alternative ρ
$\mathbf{v}$	$\mathbf{v} = \begin{bmatrix} v_1 & \cdots & v_\rho & \cdots \end{bmatrix}^T$
$\mathbf{s}$	Information signal correspond to $\mathbf{v}$
$F(\cdot)$	Distribution that denotes the information strategy of the DM
$p(\cdot)$	Distribution that denotes the prior knowledge of the DM
$A(\cdot)$	Function that denotes the maximized expected utility payoff corresponding to a posterior knowledge
$a(\cdot)$	Function that denotes the decision result of alternative decision corresponding to a posterior knowledge
$U(\cdot)$	Function that denotes the ex-ante expected utility payoff corresponding to an information strategy
$H(\cdot)$	Function that denotes the Shannon entropy corresponding to a probability distribution
$E_H(\cdot)$	Function that denotes the expected Shannon entropy of posterior knowledge corresponding to an information strategy
$\hat{c}(\cdot)$	Function that denotes the information cost corresponding to an information strategy
$I(\cdot)$	Function that denotes the decision target of the information strategy
$P(\rho \mid \mathbf{v})$	Function that denotes the probability of the DM choose alternative ρ conditional on $\mathbf{v}$
$P(\rho)$	Function that denotes the unconditional probability of the DM choose ρ
x_ρ^χ	χ -th variate of alternative ρ that influence the DM's decision, $\chi = 1, 2, \dots,$
$\mathbf{x}^\chi$	$\mathbf{x}^\chi = [x_1^\chi, \dots, x_\rho^\chi, \dots]^T$
$\hat{\mathbf{x}}$	$\hat{\mathbf{x}} = [\mathbf{x}^1, \dots \mathbf{x}^\chi, \dots]$
β^χ	Linear utility payoff coefficient corresponding to $\mathbf{x}^\chi$
$\boldsymbol{\beta}$	$\boldsymbol{\beta} = \begin{bmatrix} \beta^1 & \cdots & \beta^\chi & \cdots \end{bmatrix}^T$
$\mathbf{s}^\chi$, $\hat{\mathbf{s}}$	Information signal corresponding to $\mathbf{x}^\chi$ and $\hat{\mathbf{x}}$
$\hat{\mathbf{x}}^{\tilde{\chi}}$	$\hat{\mathbf{x}}^{\tilde{\chi}} = \begin{bmatrix} \mathbf{x}^1 & \cdots & \mathbf{x}^{\chi-1} & \mathbf{x}^{\chi+1} & \cdots \end{bmatrix}$
$P(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}})$	Function that denotes the partial conditional probabilities that conditional on $\hat{\mathbf{x}}^{\tilde{\chi}}$

Appendix B: Proof for LEMMA 1

LEMMA 1 can be proved by contradiction, considering its converse proposition can be proved false easily.

For the optimal strategy (F, a), assuming there exist an alternative ρ such that $P(\rho) > 0$, and there exist S_ρ^1, S_ρ^2 satisfy the following condition: $\left(\exists\ S_\rho^1 \cup S_\rho^2 = S_\rho\right) \wedge \left(S_\rho^1 \cap S_\rho^2 = \varnothing\right)$ $\wedge\left(\iint_{\hat{\mathbf{s}}\in S_\rho^k} F\left(\hat{\mathbf{x}}, \hat{\mathbf{s}}\right) d\hat{\mathbf{s}} d\hat{\mathbf{x}} \neq 0, k = 1, 2\right) \wedge \left(\forall\left(\hat{\mathbf{s}}^1 \in S_\rho^1 \wedge \hat{\mathbf{s}}^2 \in S_\rho^2\right), F\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}^1\right) \neq F\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}^2\right)\right)$.

Based on S_ρ^1, S_ρ^2 we can construct another feasible strategy $\bar{F}$, which can generate the same expected payoff from the strategy F and at a lower information-processing cost. The new strategy $\bar{F}$ is generated from the original strategy by relocating the probability mass from $S_\rho^1 \cup S_\rho^2$,. Intuitively, we are scrambling the signal so that the DM does not observe signals in S_ρ^1 or S_ρ^2 separately, but just observes them as one signal, denoted by $\hat{\mathbf{s}}$. Thus, the following equations are satisfied.

$$\int_{\hat{\mathbf{s}}\in S_\rho} F\left(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}\right) d\hat{\mathbf{s}} = \int_{\hat{\mathbf{s}}'\in S_\rho} \bar{F}\left(\hat{\mathbf{s}}' \mid \hat{\mathbf{x}}\right) d\hat{\mathbf{s}}' \tag{30}$$

$$\int_{\hat{\mathbf{s}}\in S_\rho} F\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}\right) F\left(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}'\right) d\hat{\mathbf{s}} = \int_{\hat{\mathbf{s}}'\in S_\rho} \bar{F}\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}'\right) \bar{F}\left(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}'\right) d\hat{\mathbf{s}}' \tag{31}$$

Thus, the following Eq. holds:

$$\begin{aligned} &\int_{\mathbf{s}\in S_\rho^1 \cup S_\rho^2} \int v_\rho F\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}\right) d\hat{\mathbf{x}} F\left(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}'\right) d\hat{\mathbf{s}} \\ &= \int_{\hat{\mathbf{s}}'\in S_\rho} \int v_\rho \bar{F}\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}'\right) d\hat{\mathbf{x}} \bar{F}\left(\hat{\mathbf{s}}' \mid \hat{\mathbf{x}}'\right) d\hat{\mathbf{s}}' \end{aligned}, \forall \rho. \tag{32}$$

The new action strategy satisfies $\bar{a}\left(\bar{F}\left(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}_\rho\right)\right) = \rho$ by the law of iterated expectations:

$$\begin{aligned} &E\left(v_\rho \mid \hat{\mathbf{s}}'\right) = E\left(E\left(v_\rho \mid \hat{\mathbf{s}} \in S_\rho\right) \mid \hat{\mathbf{s}}'\right) \\ &> E\left(E\left(v_{\rho'} \mid \hat{\mathbf{s}} \in S_\rho\right) \mid \hat{\mathbf{s}}'\right) = E\left(v_{\rho'} \mid \hat{\mathbf{s}}'\right) \end{aligned}, \rho \neq \rho'. \tag{33}$$

Hence, Eq. (34) holds.

$$
\begin{aligned}
&\int_{\hat{\mathbf{s}}\in \mathrm{S}_\rho^1 \cup \mathrm{S}_\rho^2} \max_{\forall \rho} \int U(\rho \mid \hat{\mathbf{x}}) F(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}) d\hat{\mathbf{x}} F(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}') d\hat{\mathbf{s}} \\
&= \int_{\hat{\mathbf{s}}' \in \mathrm{S}_\rho} \max_{\forall \rho} \int U(\rho \mid \hat{\mathbf{x}}) F(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}') d\hat{\mathbf{x}} F(\hat{\mathbf{s}}' \mid \hat{\mathbf{x}}') d\hat{\mathbf{s}}'
\end{aligned} \tag{34}
$$

The expected payoff of constructed strategy $(\bar{F}, \bar{a})$ can be written as:

$$
\begin{aligned}
&\iint_{\hat{\mathbf{s}}\notin \mathrm{S}_\rho^1 \cup \mathrm{S}_\rho^2} \max_{\forall \rho} \int v_\rho \bar{F}(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}) d\mathbf{X} \bar{F}(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}') d\hat{\mathbf{s}}\, p(\hat{\mathbf{x}}') d\hat{\mathbf{x}}' \\
&+ \iint_{\hat{\mathbf{s}}' \in \mathrm{S}_\rho} \max_{\forall \rho} \int v_\rho \bar{F}(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}') d\hat{\mathbf{x}} \bar{F}(\hat{\mathbf{s}}' \mid \hat{\mathbf{x}}') d\hat{\mathbf{s}}' p(\hat{\mathbf{x}}') d\hat{\mathbf{x}}'
\end{aligned}, \tag{35}
$$

there, the first term is unaffected by the change of strategy from F to $\bar{F}$, as $F(\hat{\mathbf{x}} \mid \hat{\mathbf{s}}) = \bar{F}(\hat{\mathbf{x}} \mid \hat{\mathbf{s}})$ and $F(\hat{\mathbf{s}} \mid \hat{\mathbf{x}}) = \bar{F}(\hat{\mathbf{s}} \mid \hat{\mathbf{x}})$ when $\hat{\mathbf{s}} \notin \mathrm{S}_\rho^1 \cup \mathrm{S}_\rho^2$. The second term is also unaffected because of Eq. (34). Therefore, the new strategy $\bar{F}$ can generate the same expected payoff from the strategy F.

As the entropy is a concave function of the distribution (Cover, 1999), and when $\hat{\mathbf{s}} \in \mathrm{S}_\rho^1 \cup \mathrm{S}_\rho^2$ Eq. (31) holds, the cost of information for new strategy $\bar{F}$ is lower than the original strategy F. Thus the new strategy $\bar{F}$ can generate the same expected payoff from the strategy F and at a lower information-processing cost. This means that strategy F is not the optimal strategy, and the assumption that under the optimal strategy (F, a), the posterior knowledge led by different signals corresponding to the same alternative decision may different does not hold, meaning that LAMMA 1 holds.

Appendix C: Derivation of Eq. (16)

To derive Eq. (16) we use the implication from LEMMA 1 that $F\left(\hat{\mathbf{x}}\,|\,\hat{\mathbf{s}}\right)$ are constant for all $\mathbf{s} \in \mathrm{S}_{\rho}$. The last step of Eq. (36) uses the relationship $P\left(X\,|\,Y\right)P\left(Y\right) = P\left(Y\,|\,X\right)P\left(X\right)$.

$$\begin{aligned}
U\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)\right) &= \int\left(\int A\left(F\left(\cdot\,|\,\hat{\mathbf{s}}\right)\right)F\left(\hat{\mathbf{s}}\,|\,\hat{\mathbf{x}}\right)d\hat{\mathbf{s}}\right)p\left(\hat{\mathbf{x}}\right)d\hat{\mathbf{x}} \\
&= \int\left(\int\left(\max_{\forall\rho}\int\left(v_{\rho}F\left(\hat{\mathbf{x}}\,|\,\hat{\mathbf{s}}\right)\right)d\hat{\mathbf{x}}F\left(\hat{\mathbf{s}}\,|\,\hat{\mathbf{x}}'\right)\right)d\hat{\mathbf{s}}\,p\left(\hat{\mathbf{x}}'\right)\right)d\hat{\mathbf{x}}' \\
&= \sum_{\forall\rho}\int v_{\rho}F\left(\hat{\mathbf{x}}\,|\,\mathrm{S}_{\rho}\right)d\hat{\mathbf{x}}\int\int_{\hat{\mathbf{s}}\in\mathrm{S}_{\rho}}F\left(\hat{\mathbf{s}}\,|\,\hat{\mathbf{x}}'\right)d\hat{\mathbf{s}}\,p\left(\hat{\mathbf{x}}'\right)d\hat{\mathbf{x}}' \\
&= \sum_{\forall\rho}\int v_{\rho}F\left(\hat{\mathbf{x}}\,|\,\mathrm{S}_{\rho}\right)P\left(\rho\right)d\hat{\mathbf{x}} \\
&= \sum_{\forall\rho}\int v_{\rho}P\left(\rho\,|\,\hat{\mathbf{x}}\right)p(\hat{\mathbf{x}})d\hat{\mathbf{x}}
\end{aligned} \tag{36}$$

Appendix D: Derivation of Eq. (18)

According to: (1) the additivity of mutual information, (2) the symmetry of mutual information, $H(X)-E_Y(H(X\,|\,Y))=H(Y)-E_X(H(Y\,|\,X))$, and (3) the joint distribution for the signal and state is the same as that of the action and state (LEMMA 1, one posterior leads to each action), the difference (mutual information) between Shannon entropy of prior knowledge and expected Shannon entropy of posterior knowledge for each variate can be transformed as:

$$\begin{aligned}
&H\left(p\left(\mathbf{x}^{\chi}\right)\right)-E_H\left(F\left(\mathbf{x}^{\chi},\mathbf{s}^{\chi}\right)\right)\\
&=E_H\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}^{\tilde{\chi}}\right)\right)-E_H\left(F\left(\hat{\mathbf{x}}^{\tilde{\chi}},\hat{\mathbf{s}}^{\tilde{\chi}}\right)\right)-E_H\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)\right)+E_H\left(F\left(\hat{\mathbf{x}}^{\tilde{\chi}},\hat{\mathbf{s}}^{\tilde{\chi}}\right)\right)\\
&=E_H\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}^{\tilde{\chi}}\right)\right)-E_H\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)\right)\\
&=H\left(\mathbf{s}^{\chi}\right)-E_H\left(F\left(\mathbf{s}^{\chi}\,|\,\hat{\mathbf{x}}\right)\right)\\
&=H\left(\mathbf{s}^{\chi}\right)-E_H\left(F\left(\mathbf{s}^{\chi}\,|\,\mathbf{x}^{\chi}\right)\right)\\
&=\int H\left(P\left(\rho\,|\,\hat{\mathbf{x}}\right)\right)p\left(\hat{\mathbf{x}}\right)d\hat{\mathbf{x}}-\int H\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)p\left(\hat{\mathbf{x}}^{\tilde{\chi}}\right)d\hat{\mathbf{x}}^{\tilde{\chi}}
\end{aligned}\quad. \tag{37}$$

Appendix E: Derivation of Eq. (22)

We now present the derivation of Eq. (22). The Lagrangian of the problem can be formulated as

$$U\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)\right)-\hat{c}\left(F\left(\hat{\mathbf{x}},\hat{\mathbf{s}}\right)\right)+\int\sum_{\forall\rho}\xi_{\rho}\left(\hat{\mathbf{x}}\right)P\left(\rho\,|\,\hat{\mathbf{x}}\right)p\left(\hat{\mathbf{x}}\right)d\hat{\mathbf{x}}-\int\mu\left(\hat{\mathbf{x}}\right)\left(\sum_{\forall\rho}P\left(\rho\,|\,\hat{\mathbf{x}}\right)-1\right)p\left(\hat{\mathbf{x}}\right)d\hat{\mathbf{x}} \quad (38)$$

where $\xi_{\rho}\left(\hat{\mathbf{x}}\right)$ is the Lagrange multiplier associated with Eq. (20), and $\mu\left(\hat{\mathbf{x}}\right)$ is the multiplier associated with Eq. (21). The first-order condition concerning $P\left(\rho\,|\,\hat{\mathbf{x}}\right)$ is:

$$v_{\rho}+\xi_{\rho}\left(\hat{\mathbf{x}}\right)-\mu\left(\hat{\mathbf{x}}\right)+\sum_{\forall\chi}\lambda^{\chi}\left(\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)+1-\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}\right)\right)-1\right)=0\,. \quad (39)$$

Eq. (39) implies that if $P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)>0$ and $v_{\rho}\neq-\infty$, then Eq. (20) holds. Then Eq. (39) can be transformed as:

$$\begin{gathered}\sum_{\forall\chi}\lambda^{\chi}\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}\right)\right)=v_{\rho}-\mu\left(\hat{\mathbf{x}}\right)+\sum_{\forall\chi}\lambda^{\chi}\left(\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)\right)\\ \log\left(P\left(\rho\,|\,\hat{\mathbf{x}}\right)\right)=\frac{v_{\rho}-\mu\left(\hat{\mathbf{x}}\right)+\sum_{\forall\chi}\lambda^{\chi}\left(\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)\right)}{\lambda'}\\ P\left(\rho\,|\,\hat{\mathbf{x}}\right)=\exp\left(\frac{v_{\rho}-\mu\left(\hat{\mathbf{x}}\right)}{\lambda'}\right)\left(\prod_{\forall\chi}\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)^{\lambda^{\chi}}\right)^{\frac{1}{\lambda'}}\end{gathered}\,. \quad (40)$$

Substituting Eq. (40) into Eq. (21) we can obtain:

$$\begin{gathered}1=\sum_{\forall\rho}\left(\exp\left(\frac{v_{\rho}-\mu\left(\hat{\mathbf{x}}\right)}{\lambda'}\right)\left(\prod_{\forall\chi}\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)^{\lambda^{\chi}}\right)^{\frac{1}{\lambda'}}\right)\\ \Rightarrow 1=\sum_{\forall\rho}\left(\left(\frac{\exp\left(v_{\rho}\right)}{\exp\left(\mu\left(\hat{\mathbf{x}}\right)\right)}\right)^{\frac{1}{\lambda'}}\left(\prod_{\forall\chi}\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)^{\lambda^{\chi}}\right)^{\frac{1}{\lambda'}}\right),\\ \Rightarrow e^{\mu(\hat{\mathbf{x}})/\lambda'}=\sum_{\forall\rho}\left(e^{v_{\rho}/\lambda'}\prod_{\forall\chi}\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)^{\lambda^{\chi}/\lambda'}\right)\end{gathered} \quad (41)$$

where $\lambda=\sum_{\forall\chi}\lambda^{\chi}$. By plugging Eq. (41) to Eq. (40), we can obtain Eq (22).

Appendix F:Proof for Eq. (23)

$P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)$ is defined as the partial conditional probability of choosing route ρ conditional on $\hat{\mathbf{x}}^{\tilde{\chi}}$ is determined, which means any determined $\hat{\mathbf{x}}^{\tilde{\chi}}$ corresponds to a $P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)$. The objective function represented by Eq. (19) can be transformed as follows:

$$\begin{aligned}
&\sum_{\forall\rho}\int v_\rho P(\rho \mid \hat{\mathbf{x}})\, p(\hat{\mathbf{x}})\, d\hat{\mathbf{x}} \\
&-\sum_{\forall\chi}\lambda^{\chi}\left(\begin{aligned}&-\int\sum_{\forall\rho}P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)\log\left(P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)p\left(\hat{\mathbf{x}}^{\tilde{\chi}}\right)d\hat{\mathbf{x}}^{\tilde{\chi}}\\ &+\int\sum_{\forall\rho}P(\rho \mid \hat{\mathbf{x}})\log\left(P(\rho \mid \hat{\mathbf{x}})\right)p(\hat{\mathbf{x}})\,d\hat{\mathbf{x}}\end{aligned}\right). \\
&=\sum_{\forall\rho}\int P(\rho \mid \hat{\mathbf{x}})\left(v_\rho-\lambda'\log\left(P(\rho \mid \hat{\mathbf{x}})\right)\right)p(\hat{\mathbf{x}})\,d\hat{\mathbf{x}} \\
&+\sum_{\forall\chi}\lambda^{\chi}\left(\int\sum_{\rho}P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)\log\left(P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)p\left(\hat{\mathbf{x}}^{\tilde{\chi}}\right)d\hat{\mathbf{x}}^{\tilde{\chi}}\right)
\end{aligned} \tag{42}$$

The parentheses part in the integral term of the first term in Eq. (42) can be transformed by plugging the choice probability represented by Eq. (22) as:

$$\begin{aligned}
&v_\rho-\lambda'\log\left(P(\rho \mid \hat{\mathbf{x}})\right) \\
&=v_\rho-\lambda'\log\left(\frac{\exp\left(v_\rho/\lambda'\right)\cdot\prod_{\forall\chi}P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}}{\sum_{\forall\rho}\left(\exp\left(v_\rho/\lambda'\right)\cdot\prod_{\forall\chi}P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)}\right) \\
&=-\lambda'\left(\log\left(\prod_{\forall\chi}P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)-\log\left(\sum_{\forall\rho}\left(\exp\left(v_\rho/\lambda'\right)\cdot\prod_{\forall\chi}P\left(\rho \mid \hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)\right)\right)
\end{aligned} \tag{43}$$

Substituting Eq. (43) into Eq. (42), Eq. (42) can be transformed to:

$$\begin{aligned}&\sum_{\forall\rho}-\lambda'\int\left(\sum_{\forall\chi}\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)\right)P(\rho\,|\,\hat{\mathbf{x}})\,p(\hat{\mathbf{x}})d\hat{\mathbf{x}}\\&+\sum_{\forall\chi}\lambda^{\chi}\left(\int\sum_{\forall\rho}\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)p\left(\hat{\mathbf{x}}^{\tilde{\chi}}\right)d\hat{\mathbf{x}}^{\tilde{\chi}}\right)\qquad,\\&+\lambda'\sum_{\forall\rho}\int P(\rho\,|\,\hat{\mathbf{x}})\log\left(\sum_{\forall\rho}\left(e^{v_{\rho}/\lambda'}\cdot\prod_{\forall\chi}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)\right)p(\hat{\mathbf{x}})d\hat{\mathbf{x}}\end{aligned}\tag{44}$$

where the first term can be transformed to :

$$\begin{aligned}&\sum_{\forall\rho}-\lambda'\int\left(\sum_{\forall\chi}\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)\right)P(\rho\,|\,\hat{\mathbf{x}})\,p(\hat{\mathbf{x}})d\hat{\mathbf{x}}\\&=-\sum_{\forall\chi}\lambda^{\chi}\int\sum_{\forall\rho}\log\left(P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)\right)P(\rho\,|\,\hat{\mathbf{x}})\,p(\hat{\mathbf{x}})d\hat{\mathbf{x}}\end{aligned}\ .\tag{45}$$

Eq. (45) is the opposite of the second term in Eq. (44). Thus, Eq. (42) equals Eq. (46); the optimization problem in sub-section 3.3.2 can be transformed as maximizing Eq. (46) constrained by Eq. (47).

$$\lambda'\int\log\left(\sum_{\forall\rho}\left(e^{v_{\rho}/\lambda'}\cdot\prod_{\forall\chi}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)\right)p(\hat{\mathbf{x}})d\hat{\mathbf{x}}\,,\tag{46}$$

$$\sum_{\forall\rho}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)=1\,.\tag{47}$$

When $\hat{\mathbf{x}}^{\tilde{\chi}}$ is determined, maximize Eq. (46) with $P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)$ is equivalent to maximize Eq. (48).

$$\lambda'\int\log\left(\sum_{\forall\rho}\left(e^{v_{\rho}/\lambda'}\cdot\prod_{\forall\chi}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)\right)p\left(\mathbf{x}^{\chi}\right)d\mathbf{x}^{\chi}\ .\tag{48}$$

The Lagrangian of the problem can be formulated as

$$\max\ \lambda'\int\log\left(\sum_{\forall\rho}\left(e^{v_{\rho}/\lambda'}\cdot\prod_{\forall\chi}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)\right)p\left(\mathbf{x}^{\chi}\right)d\mathbf{x}^{\chi}-\sum_{\forall\chi}\mu^{\chi}\left(\sum_{\forall\rho}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)-1\right).\tag{49}$$

Here, μ^{χ} is the multiplier associated with Eq. (47). The first-order condition with respect to $P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)$ is:

$$\lambda^{\chi}\int\frac{e^{v_\rho/\lambda'}\cdot\prod_{\chi'\neq\chi}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}'}\right)^{\lambda^{\chi'}/\lambda'}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'-1}}{\sum_{\forall\rho}\left(e^{v_\rho/\lambda'}\cdot\prod_{\forall\chi}P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)^{\lambda^{\chi}/\lambda'}\right)}p\left(\mathbf{x}^{\chi}\right)d\mathbf{x}^{\chi}=\mu^{\chi}\,. \tag{50}$$

Multiplying by $P\left(\rho\,|\,\hat{\mathbf{x}}^{\tilde{\chi}}\right)$ to both sides of Eq. (50) gives: $\lambda^{\chi}=\mu^{\chi}$. Thereby, Eq. (23) holds.

Appendix G: Notation in Section 4

$\mathrm{L}_{n\rho}$	An ordered set of lines for passenger-*n* travels via route ρ
$l_{n\rho\tau}$	τ -th line used by passenger-*n* travels via route ρ , $l_{n\rho\tau} \in \mathrm{L}_{n\rho}$, $\tau = 1, 2, \ldots, \|\mathrm{L}_{n\rho}\|$
$x_{n\rho}^{\mathrm{trans}}$	Number of transfer time when passenger-*n* travel via route ρ , $x_{n\rho}^{trans} = \left\|\mathrm{L}_{n\rho}\right\|$-1
$x_{n\rho\tau}^{\mathrm{wait}}$, $x_{n\rho\tau}^{\mathrm{walk}}$	Waiting and walking time before boarding line $l_{n\rho\tau}$
$\mathrm{K}_{n\rho}$	An order set of line sections passenger-*n* will traverse on line $l_{n\rho\tau}$
$k_{n\rho\kappa}$	κ -th line section used by passenger-*n* travels via route ρ , $k_{n\rho\kappa} \in \mathrm{K}_{n\rho}$, $\kappa = 1, 2, \ldots \|\mathrm{K}_{n\rho}\|$
$x_{n\rho\kappa}^{\mathrm{travel}}$ $x_{n\rho\kappa}^{\mathrm{stand}}$ $x_{n\rho\kappa}^{\mathrm{crowd}}$	Expected travel, standing time and crowding level in section $k_{n\rho\kappa}$
$D_{n\rho\kappa}$	Standing passenger density on section $k_{n\rho\kappa}$
$x_{n\rho}^{\mathrm{trans}}$ $x_{n\rho}^{\mathrm{wait}}$ $x_{n\rho}^{\mathrm{walk}}$ $x_{n\rho}^{\mathrm{travel}}$ $x_{n\rho}^{\mathrm{stand}}$ $x_{n\rho}^{\mathrm{crowd}}$	Number of transfer times, waiting time, walking time, travel time, standing time, and crowding level when passenger-*n* travel via route ρ
$\mathbf{x}_n^{\mathrm{trans}}$	$\mathbf{x}_n^{trans} = \left[x_{n1}^{trans} \quad \cdots \quad x_{n1}^{trans} \quad \cdots \right]^T$, similar for $\mathbf{x}_n^{\mathrm{wait}}$ $\mathbf{x}_n^{\mathrm{walk}}$ $\mathbf{x}_n^{\mathrm{travel}}$ $\mathbf{x}_n^{\mathrm{stand}}$ $\mathbf{x}_n^{\mathrm{crowd}}$
$\hat{\mathbf{x}}_n$	$\hat{\mathbf{x}}_n = \left[\mathbf{x}_n^{\mathrm{trans}} \quad \mathbf{x}_n^{\mathrm{wait}} \quad \mathbf{x}_n^{\mathrm{walk}} \quad \mathbf{x}_n^{\mathrm{travel}} \quad \mathbf{x}_n^{\mathrm{stand}} \quad \mathbf{x}_n^{\mathrm{crowd}} \right]$
$\boldsymbol{\beta}$	Set of coefficients for calculating utility, $\boldsymbol{\beta} = \left\{ \beta^{\mathrm{trans}}, \beta^{\mathrm{walk}}, \beta^{\mathrm{wait}}, \beta^{\mathrm{travel}}, \beta^{\mathrm{stand}}, \beta^{\mathrm{crowd}} \right\}$
$\boldsymbol{\lambda}$	Set of basic unit information cost for different variates, $\boldsymbol{\lambda} = \left\{ \lambda^{\mathrm{trans}}, \lambda^{\mathrm{walk}}, \lambda^{\mathrm{wait}}, \lambda^{\mathrm{travel}}, \lambda^{\mathrm{stand}}, \lambda^{\mathrm{crowd}} \right\}$
$\boldsymbol{\lambda}_n$	Set of unit information cost for different variates of passenger-*n* $\lambda_n = \left\{ \lambda_n^{trans}, \lambda_n^{walk}, \lambda_n^{wait}, \lambda_n^{travel}, \lambda_n^{stand}, \lambda_n^{crowd} \right\}$
$y_{n\rho}$	$y_{n\rho} = 1$ if passenger-*n* choose route ρ and zero otherwise.