UNIVERSITY OF LEEDS

This is a repository copy of *A Generalized Rationally Inattentive Route Choice Model with Non-uniform Marginal Information Costs*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/213021/

Version: Accepted Version

Article:

Zhou, B. and Liu, R. orcid.org/0000-0003-0627-3184 (2024) A Generalized Rationally Inattentive Route Choice Model with Non-uniform Marginal Information Costs. Transportation Research Part B: Methodological, 189. 102993. ISSN 0191-2615

https://doi.org/10.1016/j.trb.2024.102993

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

25th International Symposium on Transportation and Traffic Theory, ISTTT 25, 15-17 July 2024
Ann Arbor, Michigan, United States

A Generalized Rationally Inattentive Route Choice Model with Non-uniform Marginal Information Costs

Bo Zhou[1] and Ronghui Liu[*]

Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, United Kingdom

Abstract

Information consumes attention. In the information-rich society, a wealth of information can support decision-making, but it can also create poverty of attention to and inability to make the best use of information. This study applies the theory of rational inattention in modelling traveller's route choice behaviour, where the attention (costs) to information is non-uniform. We establish the mathematical formulation of a generalized rationally inattentive route choice model with non-uniform marginal information costs, and prove that the optimal conditional route choice probabilities for all the routes always locate within the interior of the feasible region of the route choice model. Based on this property, we analytically characterize the closed-form expression of the optimal conditional choice probabilities, and devise an efficient iterative solution algorithm to compute them. Finally, two numerical examples are conducted to demonstrate the theoretical properties of the rationally inattentive route choice behaviour. This behavioural modelling approach provides an insight on how the rationally inattentive travellers spontaneously learn the optimal route choice from the acquired information.

Keywords: Route choice; Rational inattention; Non-uniform marginal information costs; Closed-form expression.

1 Introduction

Route choice is a critical area of studies in travel behaviour and transportation network analysis. The development of route choice models has been influenced by advancements in random utility-based discrete choice models (Ben-Akiva and Lerman, 1985). These models typically incorporate exogenous random shocks that obey (generalized) independently and identically distributed (i.i.d) Gumbel distribution (Duncan et al., 2020; Fosgerau et al., 2013; Knies et al., 2022; Ma and Fukuda, 2015; Mai, 2016; Mai et al., 2015; Papola et al., 2018). Nowadays, in our information-rich society, the introduction of new communication and sensing techniques has changed the way that the travellers plan their journeys. For instance, the availability of smart phones, mapping software, and social media platforms has significantly increased the amount of traffic information accessible to travellers. With the aid of these communication and information technologies, travellers can now obtain real-time traffic conditions across the transportation network and use this information

* Corresponding author. E-mail address: R.Liu@its.leeds.ac.uk.
[1] The research was conducted during the first author's visit at ITS Leeds. Current address: School of Mathematics and Statistics, Chongqing Jiaotong University, China.

to inform their travel choices (Mai et al., 2021). Consequently, the driving force behind route choice behaviour has seen a gradually shift from being exogenous to endogenous, with travellers spontaneously choosing their routes based on an assessment of the acquired information on traffic conditions. This calls for novel forms of models to represent this shift in traveller's route choice behaviour.

Another behaviour emerging is that information consumes attention. Simon (1955) is among the first to establish a link between information and attention, suggesting that attention is a valuable but limited resource and a wealth of information creates a poverty of attention. In the route choice behaviour, this means that the travellers would align information with their available attentional resources to make a choice decision. In contemporary research, the rational inattention (RI) framework, initially introduced by Sims (2003, 2006, 2010), quantifies the expenditure of attention as the information cost using Shannon entropy (Shannon, 1948). Within this framework, the travellers determine what and how much information to acquire, striking a balance between the attention consumed by acquiring information and the benefits derived from having more informed expectations. Building upon these pioneering works, Matějka and Mckay (2015) bridges the gap between RI choice behaviour and discrete choice models. They find that the optimal conditional choice probabilities resemble a generalized multinomial logit (GMNL) model, which can be regarded as an extension of the standard logit choice model under the random utility modelling framework (Wang, 2021).

The RI framework has found extensive applications across various fields, including economics (Caplin et al., 2019; Fosgerau et al., 2020; Steiner et al., 2017), marketing (Boyacı and Akçay, 2018; Matějka, 2016; Ravid, 2020), as well as transportation (Fosgerau and Jiang, 2019; Habib, 2023; Jiang et al., 2020). In the realm of transportation research, Fosgerau and Jiang (2019) develop a theoretical model that features the RI traveller who aims to acquire information on the traffic conditions so as to optimally choose the departure time of their daily commuting. Expanding upon this work, Jiang et al. (2020) further extend the RI framework to address the route choice problem within a stochastic network, wherein the travellers acquire information on the traffic conditions so as to optimally choose the routes for their journeys. Habib (2023) proposes econometric formulations of rationally inattentive choice models. The proposed models are estimated for the commuting mode choices in Greater Toronto and Hamilton Area. Although Fosgerau and Jiang (2019) and Jiang et al. (2020) focus on distinct topics, their models exhibit a shared characteristic in terms of optimal conditional choice probabilities for the candidates: they allow the conditional choice probabilities for a set of candidates locate on the boundary of the feasible region. For instance, in Jiang et al. (2020), the optimal conditional choice probabilities for the dominated routes (a route is called dominated, if its travel cost is always higher than another route with respect to all the acquired information) remain zero, indicating that they always locate on the boundary. This pattern of RI choice behaviour, inherited from Matějka and Mckay (2015), is prevalent in the existing literature (Caplin et al., 2019; Huettner et al., 2019; Walker-Jones, 2023).

On the other hand, while Shannon entropy has proven to be a valuable tool, it does have limitations in route choice behaviour modelling framework as they are not what it is designed for. For instance, Shannon entropy does not allow for different levels of attention corresponding to acquiring information from different sub-components because it is a one-parameter approach for quantifying the information cost and thus can only incorporate a unique level of attention. This can restrict the effectiveness of applying Shannon entropy in practical settings (Dean and Neligh, 2023). In order to ease such a restriction, Huettner et al. (2019) and Walker-Jones (2023) provide multi-parameter approaches for quantifying the information cost. Huettner et al. (2019) quantifies the information cost across each sub-component of information, whilst Walker-Jones (2023) quantifies the information cost across each attribute of information, in which an attribute is a group of sub-components.

In this study, we aim to develop a generalized RI route choice model based on the multi-parameter approach, and to characterize the general RI route choice behaviour. This study offers three key contributions. First, we establish the mathematical formulation of a generalized RI route choice model with non-uniform marginal information costs, in which the marginal information costs are utilized to quantify the expenditure of attention on acquiring each unit information. Different to the existing RI choice models, such as Huettner et al. (2019), Jiang et al. (2020), Matějka and Mckay (2015) and Walker-Jones (2023), we introduce the concept of background information in our model, which is objectively provided by the transportation network and keeps invariant over the entire decision-making process, to describe the environment that the RI traveller faces before the information acquisition process launches. We then prove that such a generalized RI route choice model makes all the optimal conditional route choice probabilities locate within the interior of the feasible region of the route choice model. Second, we analytically characterize the closed-form expression for the optimal conditional route choice probabilities associated with all the candidate routes, which provide a more comprehensive understanding of the RI route choice behaviour. Although Huettner et al. (2019), Jiang et al. (2020), Matějka and Mckay (2015) and Walker-Jones (2023) have characterized the structure of the optimal conditional choice probabilities, their results can only be served as necessary but insufficient conditions for optimality, as such, there is no closed-form expressions of optimal conditional route choice probabilities in their works. Third, we devise an iterative solution algorithm based on the closed-form expression, which provides a more streamlined solution process by simplifying the requirements of numerous additional necessary and sufficient conditions for checking the convergence in Caplin et al. (2019), Huettner et al. (2019), Jiang et al. (2020) and Walker-Jones (2023).

2 An overview of the RI route choice model with uniform marginal information costs

2.1 A brief introduction on Shannon information theory

The original RI concept is introduced in Sims (2003), which leverages several useful instruments provided in the Shannon information theory (Shannon, 1948) to measure the amount of acquired information when making choice decisions. It starts with the Shannon entropy (entropy for short hereafter) that measures the amount of information carried by a random variable, whose unit is “bit”. Entropy measures the uncertainty of a random variable. A higher entropy indicates that the random variable carries larger amount of information, which implies a higher level of uncertainty. A lower entropy on the other hand indicates the random variable carries lower amount of information, which suggests it is more predictable. The definition of entropy applies to continuously as well as discretely distributed variables. We take the discrete random variable as an illustrative example. For a random variable X with probability mass function $p(\cdot)$, the amount of information carried by X can be measured by the unconditional entropy as:

$$\mathscr{H}(p;X) = -\sum_{x} p(x)\log p(x), \tag{1}$$

where the logarithm function can be of any base, because the base only determines a scale factor for the information measure. In practice, it is conventional to take the natural logarithm with Euler's number e. For instance, suppose X follows a discrete distribution with equal probabilities on 0 and 1, its entropy can be calculated by (1) as

$$\mathscr{H}(p;X) = -\left(\frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2}\right) = 0.693,$$

which means that random variable X carries 0.693 bit of information. In addition, if X is a deterministic variable, that is, $p(x) = 0$ or 1 for all possible values of X, then it carries no information with $\mathscr{H}(p;X) = 0$.

Based on the general definition of entropy, Shannon (1948) introduces a conditional entropy to measure the amount of information carried by one random variable from observing the other random variable(s). For instance, suppose random variable X has been obtained before acquiring any information. Then, the amount of information carried by X can be measured by $\mathscr{H}(p;X)$. Once the information on another random variable Y is acquired, the amount of information carried by X from observing Y can be measured by:

$$\mathscr{H}(p;X|Y) = -\sum_{y} g(y)\sum_{x} p(x|y)\ln p(x|y), \tag{2}$$

where $g(\cdot)$ is the probability mass function of Y and $p(\cdot|y)$ is the conditional probability mass function of X from observing Y. The conditional entropy (2) measures the expectation of the amount of information carried by the conditional distribution $\{p(x|y)\}$ with respect to Y. Based on the unconditional entropy (1) and conditional entropy (2), the amount of information acquired from Y given X can be calculated by subtracting (2) from (1)

$$\mathscr{I}(p;X,Y) = \mathscr{H}(p;X) - \mathscr{H}(p;X|Y). \tag{3}$$

For instance, suppose Y follows a discrete distribution with equal probabilities on 0 and 1, and the conditional probabilities of X given Y be $f(x=0|y=0) = 1/3$, $f(x=1|y=0) = 2/3$, $f(x=0|y=1) = 3/4$, $f(x=1|y=1) = 1/4$. Then, the conditional entropy can be calculated by (2)

$$\mathscr{H}(p;X|Y) = -\frac{1}{2}\left(\frac{1}{3}\ln\frac{1}{3} + \frac{2}{3}\ln\frac{2}{3}\right) - \frac{1}{2}\left(\frac{3}{4}\ln\frac{3}{4} + \frac{1}{4}\ln\frac{1}{4}\right) = 0.599.$$

Thus, $\mathscr{I}(X,Y)$ can be calculated by (3) as:

$$\mathscr{I}(p;X,Y) = \mathscr{H}(p;X) - \mathscr{H}(p;X|Y) = 0.094,$$

which means that the amount of acquired information on Y given X is 0.094 bit.

2.2 A brief overview of the RI route choice formulation in Jiang et al. (2020)

Sims (2003) introduced the concept of marginal information cost to quantify the expenditure of attention on acquiring each unit information. Most of the existing RI choice models assume a uniform marginal information cost across each sub-component of information. In this section, we provide a brief overview of the RI route choice formulation with uniform marginal information costs, following Jiang et al. (2020).

We consider an individual RI traveller who intends to travel through a transportation network. We denote $\mathcal{H}$ be the set of candidate routes, indexed by i. We denote a vector $\boldsymbol{\omega}$ be the state of traffic conditions (state for short hereafter), which belongs to a finite set $\boldsymbol{\Omega}$. The RI traveller has his own prior belief $g(\cdot) \in \mathscr{P}(\boldsymbol{\Omega})$ on the states, where $\mathscr{P}(\boldsymbol{\Omega})$ denotes the set of probability distributions on $\boldsymbol{\Omega}$. In practice, the RI traveller acquires information on the states with the goal of improving his route choice behaviour that are presented by the discrete probabilities $\boldsymbol{p} = \{p(i|\boldsymbol{\omega})\}$ conditioned on the states. We denote the travel cost for route i in state $\boldsymbol{\omega}$ be $c(i,\boldsymbol{\omega})$ with $0 < c(i,\boldsymbol{\omega}) < +\infty$. Then, the expected travel cost is defined as

$$C_{\text{travel}}(\boldsymbol{p}) = \sum_{i \in \mathcal{H}} \sum_{\boldsymbol{\omega} \in \Omega} c\,(i,\boldsymbol{\omega})\, g(\boldsymbol{\omega})\, p(i|\boldsymbol{\omega}), \tag{4}$$

which is a standard definition in the work on the RI route choice behaviour (Jiang et al., 2020).

Before acquiring any information, the RI route choice behaviour is independent of the states and can be presented by the unconditional route choice probabilities $\{p(i)\}$. By the law of total probability, the unconditional route choice probabilities can be expressed by the expectation of the conditional route choice probabilities with respect to the states, that is,

$$p(i) = \sum_{\boldsymbol{\omega} \in \Omega} g(\boldsymbol{\omega})\, p\,(i|\boldsymbol{\omega}), \tag{5}$$

in which the unconditional route choice probabilities can be regarded as the habitual route choice behaviour of the RI traveller.

Then, according to (1), the amount of information possessed by the RI traveller can be measured by the unconditional entropy

$$\mathscr{H}(\boldsymbol{p}) = -\sum_{i \in \mathcal{H}} p\,(i) \ln p\,(i). \tag{6}$$

Once the information on the states is acquired, the RI traveller formulates the RI route choice behaviour described by $\{p(i|\boldsymbol{\omega})\}$. Thus, according to (2), the amount of information possessed by the RI traveller can be measured by the conditional entropy

$$\mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}) = -\sum_{\boldsymbol{\omega} \in \boldsymbol{\Omega}} g\,(\boldsymbol{\omega}) \sum_{i \in \mathcal{H}} p\,(i|\boldsymbol{\omega}) \ln p\,(i|\boldsymbol{\omega}). \tag{7}$$

Accordingly, the total amount of information acquired on the states can be measured according to (3)

$$\mathscr{I}(\boldsymbol{p};\boldsymbol{\Omega}) = \mathscr{H}(\boldsymbol{p}) - \mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}). \tag{8}$$

Consequently, the information cost can be quantified using (8) as

$$C_{\text{uni,info}}(\boldsymbol{p}) = \lambda \mathscr{I}(\boldsymbol{p};\boldsymbol{\Omega}), \tag{9}$$

where $\lambda > 0$ is the marginal information cost.

With the definition of the expected travel cost (4) and the information acquisition cost (9), the RI route choice behaviour with uniform marginal information costs can then be described by the following optimization model:

$$\underset{\{p(i|\boldsymbol{\omega})\}}{\text{Minimize}}\ C_{\text{uni}}(\boldsymbol{p}) = \sum_{i\in\mathcal{H}}\sum_{\boldsymbol{\omega}\in\boldsymbol{\Omega}} c\,(i,\boldsymbol{\omega})\,p(i|\boldsymbol{\omega})\,g(\boldsymbol{\omega}) + \lambda\mathscr{I}(\boldsymbol{p};\boldsymbol{\Omega}), \tag{10}$$

$$\text{Subject to } 1 = \sum_{i\in\mathcal{H}} p\,(i|\boldsymbol{\omega}), \text{ for all } \boldsymbol{\omega}\in\boldsymbol{\Omega}, \tag{11}$$

$$0 \le p(i|\boldsymbol{\omega}), \text{ for all } i\in\mathcal{H},\ \boldsymbol{\omega}\in\boldsymbol{\Omega}, \tag{12}$$

where the RI traveller aims to find the optimal conditional route choice probabilities that minimize the sum of the travel cost and information cost. Constraint (11) ensures that at least one route should be chosen in any given state, and Constraint (12) ensures the conditional choice probabilities are non-negative in all states.

RI route choice model is related to the bounded rationality choice model and the random utility choice model, while it is quite different to the prospect choice model. In the bounded rationality choice model, the route travel costs are disturbed by the indifference parameters, which can be either random or deterministic. When the indifference parameters are random variables, which follow the i.i.d. Gumbel distribution, the bounded rationality choice behaviour can be described by the random utility choice model. When the indifference parameters are defined by the information cost, the bounded rationality choice behaviour can be described by the RI choice model. On the other hand, the prospect choice model allows the individuals learn a choice behaviour which does not necessarily minimize the total (expected) cost, because they place other considerations above utility, called travel prospects.

As discussed in Jiang et al. (2020), the structure of the optimal solutions of route choice model (10)-(12) can be characterized in the following proposition.

Proposition 1. *If $\lambda > 0$, then the solution of route choice model (10)-(12) is optimal only if*

$$p(i|\boldsymbol{\omega}) = \frac{p(i)e^{-\frac{c(i,\boldsymbol{\omega})}{\lambda}}}{\sum_{j\in\mathcal{H}} p\,(j)e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda}}}, \tag{13}$$

for all $i\in\mathcal{H}$ and $\boldsymbol{\omega}\in\boldsymbol{\Omega}$.

Condition (13) provides a GMNL formulation of the optimal conditional route choice probabilities. In the extreme case of $\lambda = 0$ where the RI traveller can acquire information on the states for free, they will make route choice decisions with complete rationality, i.e., they always choose the routes with lowest travel costs. Conversely, in the other extreme case of $\lambda = +\infty$ where information is too costly to acquire, the RI traveller would not acquire any information on the states and would choose the routes with lowest expected travel cost only according to the habitual route choice behaviour $\{p(i)\}$ (Matějka and Mckay, 2015).

Important to note that Condition (13) is necessary but insufficient for optimality, since the

Karush-Kuhn-Tucker (KKT) conditions can only be applied to the conditional route choice probabilities that locate within the interior of the feasible region of the choice model. Whereas route choice model (10)-(12) allows the optimal conditional route choice probabilities for the dominated routes to be located on the boundary. Caplin et al. (2019) has summarized the key limitation of Condition (13), that is, it can determine the RI choice behaviour among the routes that are chosen with positive probabilities, but it cannot identify which route is with positive choice probability. The existence of the zero-choice-probability routes is the major reason that results in the non-existence of the closed-form expression of the optimal conditional route choice probabilities. We postulate that the existence of zero-choice-probability routes is caused by the freely available habitual information $\{p(i)\}$, based on which the information acquisition process in the existing RI choice modelling framework has been built. The habitual route choice behaviour $p(i)$ was used to represent the choice behaviour before acquiring any information on the states in the RI choice models, but it is presented as a linear combination of the RI route choice behaviour $p(i|\boldsymbol{\omega})$ according to (5). Note that the RI route choice behaviour $p(i|\boldsymbol{\omega})$ is relevant to the full information on the states. Once the habitual information is freely available, the RI traveller can then exactly identify which route is dominated before the information acquisition process launches, and accordingly, they will not choose the dominated route. In practice, however, the habitual information is freely available only when the RI traveller is exceedingly familiar with traffic conditions, for instance, his routinely commute routes to work. Otherwise, the RI traveller should always consume attention to acquire habitual information, particularly when they intend to conduct his travel along unfamiliar routes. In this context, it is more viable to consider the scenario that the RI traveller acquires information on the states from null. Therefore, a more tenable RI choice modelling framework entails factoring in the expenditure of attention to acquire habitual information.

3 A generalized RI route choice model with non-uniform marginal information costs

3.1 Mathematical formulation of the generalized RI route choice model

In this section, we consider the scenario where acquiring information from different sub-component of states entails different levels of attention. Before the information acquisition process is launched, the RI traveller should possess null information on the states. Instead, what they face is only the background information provided by the transportation network, which keeps invariant over the entire information acquisition process. We denote the background information by $\Omega_{-1} = \mathcal{H}$. In this situation, the RI traveller only knows that there are $|\mathcal{H}|$ candidate routes for them to choose from the outset. According to the principle of maximum entropy established in Jaynes (1957), if nothing is known about a distribution except that it belongs to a certain class, then the distribution with the largest entropy should be selected as the least-informative default. Entropy maximization with no testable information only respects the universal constraint that the sum of the probabilities is one. Under this constraint, the uniform distribution is with the maximum entropy among the finite limited number of candidate routes. Thus, before launching the information acquisition process, the amount of background information can be measured by

$$\mathscr{H}(\Omega_{-1}) = -\sum_{i\in\mathcal{H}} \frac{1}{|\mathcal{H}|}\ln\frac{1}{|\mathcal{H}|} = \ln|\mathcal{H}|, \tag{14}$$

which is independent of the RI route choice behaviour.

Then, the first step to launch the information acquisition process is to acquire information from the background so as to learn the habitual information. Once the habitual information $\Omega_0 = \{p(i)\}$ is acquired, the amount of information possessed by the RI traveller can be measured by

$$\mathscr{H}(\boldsymbol{p}|\Omega_0) = -\sum_{i\in\mathcal{H}} p(i)\ln p(i). \tag{15}$$

Then, similar to (8), the amount of acquired information on the habitual information can be measured using (14) and (15) as

$$\mathscr{I}(\boldsymbol{p};\Omega_0|\Omega_{-1}) = \mathscr{H}(\Omega_{-1}) - \mathscr{H}(\boldsymbol{p}|\Omega_{-1}\times\Omega_0). \tag{16}$$

Now, we consider the process that the RI traveller acquires information on the states. We denote the state as an n-dimensional vector $\boldsymbol{\omega} = (\omega_1,\omega_2,...,\omega_n) \in \mathbb{R}^n$, which belongs to a finite set $\boldsymbol{\Omega} = \Omega_1\times\Omega_2\times\cdots\times\Omega_n$, where Ω_k is the set of possible values for the k-th sub-component of state $\boldsymbol{\omega}$. For the sake of completeness, hereafter, we denote $p(i|\boldsymbol{\omega}_{1,...,k}) = p(i)$ for $k=0$.

Before acquiring information from Ω_k with $k\geq 1$, the RI traveller has already obtained the partial state that contains the first $k-1$ sub-components of state $\boldsymbol{\omega}$, which is denoted by $\boldsymbol{\omega}_{1,...,k-1} = (\omega_1,...,\omega_{k-1})$ belongs to the set $\boldsymbol{\Omega}_{1,...,k-1} = \Omega_1\times\cdots\times\Omega_{k-1}$. The prior belief on the partial state $\boldsymbol{\omega}_{1,...,k-1}$ is given by $g(\boldsymbol{\omega}_{1,...,k-1})$. For ease of presentation, we denote $\boldsymbol{\Omega}_{-1,...,k-1} = \Omega_{-1}\times\Omega_0\times\boldsymbol{\Omega}_{1,...,k-1}$ for $k>1$ and $\boldsymbol{\Omega}_{-1,0} = \Omega_{-1}\times\Omega_0$. For the sake of completeness, we set $g(\boldsymbol{\omega}_{1,...,k-1}) = 1$ for $k=1$.

In this context, the RI route choice behaviour is presented by $\{p(i|\boldsymbol{\omega}_{1,...,k-1})\}$. Then, the amount of information possessed by the RI traveller can be measured by

$$\mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,...,k-1}) = -\sum_{i\in\mathcal{H}}\sum_{\boldsymbol{\omega}_{1,...,k-1}\in\boldsymbol{\Omega}_{1,...,k-1}} g(\boldsymbol{\omega}_{1,...k-1})\,p(i|\boldsymbol{\omega}_{1,...,k-1})\ln p(i|\boldsymbol{\omega}_{1,...,k-1}). \tag{17}$$

Once the information from Ω_k is acquired, the RI traveller obtains partial state that contains the first k sub-components $\boldsymbol{\omega}_{1,...,k} = (\omega_1,...,\omega_k)$ of state $\boldsymbol{\omega}$. The prior belief on the partial state $\boldsymbol{\omega}_{1,...,k}$ is given by $g(\boldsymbol{\omega}_{1,...,k})$. In this context, the RI route choice behaviour is presented by $\{p(i|\boldsymbol{\omega}_{1,...,k})\}$. Then, similar to (17), the amount of information possessed by the RI traveller can be measured by

$$\mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,...,k}) = -\sum_{i\in\mathcal{H}}\sum_{\boldsymbol{\omega}_{1,...,k}\in\boldsymbol{\Omega}_{1,...,k}} g(\boldsymbol{\omega}_{1,...k})\,p(i|\boldsymbol{\omega}_{1,...,k})\ln p(i|\boldsymbol{\omega}_{1,...,k}). \tag{18}$$

Accordingly, using (17) and (18), the total amount of information acquired for the k-th sub-component can be measured by

$$\mathscr{I}(\boldsymbol{p};\Omega_k|\boldsymbol{\Omega}_{-1,\dots,k-1}) = \mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,\dots,k-1}) - \mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,\dots,k}). \tag{19}$$

Consequently, the information cost can be quantified according to (16) and (19) across all the sub-components of $\boldsymbol{\Omega}$

$$C_{\text{non,info}}(\boldsymbol{p}) = \sum_{k=0}^{n} \lambda_k \mathscr{I}(\boldsymbol{p};\Omega_k|\boldsymbol{\Omega}_{-1,\dots,k-1}), \tag{20}$$

where $\lambda_k > 0$ is the marginal information cost for acquiring each unit of information from Ω_k.

Then, with the definition of the travel cost (4), the information cost (20), we formulate the following optimization model that describes the RI route choice behaviour with non-uniform marginal information costs

$$\underset{\{p(i|\boldsymbol{\omega})\}}{\text{Minimize}}\ C_{\text{non}}(\boldsymbol{p}) = \sum_{i\in\mathcal{H}}\sum_{\boldsymbol{\omega}\in\boldsymbol{\Omega}} c\,(i,\boldsymbol{\omega})\,p(i|\boldsymbol{\omega})\,g(\boldsymbol{\omega}) + \sum_{k=0}^{n} \lambda_k \mathscr{I}(\boldsymbol{p};\Omega_k|\boldsymbol{\Omega}_{-1,\dots,k-1}), \tag{21}$$

$$\text{Subject to } 1 = \sum_{i\in\mathcal{H}} p\,(i|\boldsymbol{\omega}),\ \text{ for all } \boldsymbol{\omega}\in\boldsymbol{\Omega}, \tag{22}$$

$$0 \le p(i|\boldsymbol{\omega}),\ \text{ for all } i\in\mathcal{H},\ \boldsymbol{\omega}\in\boldsymbol{\Omega}. \tag{23}$$

In the following, we make an assumption on the marginal information costs.

Assumption 1. The marginal information costs follow an ascending order with respect to the subscripts, that is, $0 < \lambda_0 < \lambda_1 < \cdots < \lambda_n < +\infty$.

In the RI route choice model (21)-(23), information acquisition process is a superposition process, that is, the amount of acquired information on one additional sub-component is derived from the entropy conditioned on the previous sub-components minus the entropy conditioned on the product of the new and the previous sub-components. For example, when the RI traveller acquires information from Ω_0 with marginal information cost λ_0 for each unit of information, the information cost is calculated by

$$\lambda_0 \mathscr{I}(\boldsymbol{p};\Omega_0|\Omega_{-1}) = \lambda_0(\mathscr{H}(\Omega_{-1}) - \mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,0})),$$

where the RI traveller has already possessed the background information Ω_{-1} and needs to acquire information from the product $\boldsymbol{\Omega}_{-1,0} = \Omega_{-1}\times\Omega_0$. When the RI traveller acquires information from Ω_1 with marginal information cost λ_1, the information cost is calculated by

$$\lambda_1 \mathscr{I}(\boldsymbol{p};\Omega_1|\boldsymbol{\Omega}_{-1,0}) = \lambda_1(\mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,0}) - \mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,0,1})),$$

where the RI traveller has already possessed the information from $\boldsymbol{\Omega}_{-1,0}$ and needs to acquire information from the product $\boldsymbol{\Omega}_{-1,0,1} = \Omega_{-1}\times\Omega_0\times\Omega_1$. Since $\boldsymbol{\Omega}_{-1,0}$ is a part of $\boldsymbol{\Omega}_{-1,0,1}$, the latter contains more information than the former. Thus, the consumption of attention on acquiring each unit information from $\boldsymbol{\Omega}_{-1,0,1}$ should be higher than that from $\boldsymbol{\Omega}_{-1,0}$. Hence, $\lambda_0 < \lambda_1$. Following the similar analogy, we can get that Assumption 1 is reasonable.

According to the principle of maximum entropy, the total amount of acquired information is always non-negative, with zero if $p(i|\boldsymbol{\omega})$ is uniformly distributed in each given state $\boldsymbol{\omega}$. When

$\lambda_0 = 0$, the RI traveller acquires the candidate information without any attentional expenditure, then RI route choice model (21)-(23) would degenerate into the choice model proposed in Huettner et al. (2019). Furthermore, on the basis of $\lambda_0 = 0$ and uniform marginal information costs for the sub-components, e.g., $\lambda_1 = \lambda_2 = \cdots = \lambda_n$, RI route choice model (21)-(23) would degenerate into the RI route choice model (10)-(12) presented in Jiang et al. (2020). Therefore, RI route choice model (21)-(23) is a generalization of the choice model considered in the existing works as Huettner et al. (2019), Jiang et al. (2020), Matějka and Mckay (2015). Naturally, a question arises: would the generalized RI route choice model (21)-(23) gives rise to a different pattern of RI route choice behaviour to those in the existing works? In the next sub-section, we will analytically characterize the RI route choice behaviour described in the route choice model (21)-(23).

3.2 Optimal route choice behaviour

The marginal information costs reflect the difficulty of acquiring information from the corresponding subcomponents: subcomponents with higher marginal information costs associated with them are more costly to learn about (Walker-Jones, 2023). Two extreme cases of marginal information costs, e.g. $\lambda_0 = \lambda_1 = 0$ and $\lambda_0 = 0, \lambda_n = +\infty$, are similar to those discussed in Huettner et al. (2019).

In order to characterize the closed-form expression of the optimal conditional route choice probabilities, the following result that ensures the positivity for the optimal solutions of route choice model (21)-(23) should be established in prior.

Theorem 1. *Let* $\{p^*(i|\boldsymbol{\omega})\}$ *be the optimal solution of route choice model (21)-(23). Under Assumption 1,* $p^*(i|\boldsymbol{\omega}) > 0$ *for all* $i \in \mathcal{H}$ *and* $\boldsymbol{\omega} \in \Omega$.

Theorem 1 guarantees that the optimal solutions of route choice model (21)-(23) locate in the interior of the feasible region of the model. This is a prerequisite to apply the KKT conditions to characterize the closed-form expressions of the optimal route choice probabilities. Theorem 1 is proven by contradiction. We start by supposing that there exists an optimal solution of route choice model (21)-(23) that assigns a zero conditional choice probability for a certain route. We then show that it is possible to construct another feasible solution that can reduce the objective function (21) by introducing a small perturbation to the zero conditional choice probability. More importantly, we find that such a small perturbation relies on a positive λ_0, which theoretically demonstrates that the route choice models without incorporating the background information fail to guarantee the optimal conditional route choice probabilities locate within the interior of the feasible region in all states. Indeed, the choice models considered in Huettner et al. (2019), Jiang et al. (2020), Matějka and Mckay (2015) and Walker-Jones (2023) allow the optimal conditional choice probabilities for the dominated alternatives locate on the boundary, i.e., the optimal choice probabilities for such alternative are always zero. Therefore, route choice model (21)-(23) describes a different pattern of RI route choice behaviour that differs from those in the existing works. The full proof is presented in Appendix A.

Now, we are ready to characterize the closed-form expression of the optimal conditional route choice probabilities.

Theorem 2. *Let* $\{p(i|\boldsymbol{\omega})\}$ *be the solutions of route choice model (21)-(23). Under Assumption 1,*

$\{p(i|\boldsymbol{\omega})\}$ *is optimal, if only if* $p(i|\boldsymbol{\omega})>0$ *and satisfies*

$$p(i|\boldsymbol{\omega})=\frac{e^{-\frac{c(i,\boldsymbol{\omega})}{\lambda_n}}\Theta(p(i|\boldsymbol{\omega}))}{\sum_{j\in\mathcal{H}}e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda_n}}\Theta(p(j|\boldsymbol{\omega}))}, \tag{24}$$

for all $i\in\mathcal{H}$ *and* $\boldsymbol{\omega}\in\boldsymbol{\Omega}$, *where* $\Theta(p(i|\boldsymbol{\omega}))=\prod_{k=1}^{n}p\left(i|\boldsymbol{\omega}_{1,\ldots,k-1}\right)^{\frac{\lambda_k-\lambda_{k-1}}{\lambda_n}}$. *The conditional route choice probability* $p(i|\boldsymbol{\omega}_{1,\ldots,k-1})$ *on partial state* $\boldsymbol{\omega}_{1,\ldots,k-1}$ *is given by*

$$p(i|\boldsymbol{\omega}_{1,\ldots,k})=\sum_{\boldsymbol{\omega}_{k+1,\ldots,n}\in\Omega_{k+1,\ldots,n}}p\left(i|\boldsymbol{\omega}_{1,\ldots,k},\boldsymbol{\omega}_{k+1,\ldots,n}\right)g\left(\boldsymbol{\omega}_{k+1,\ldots,n}|\boldsymbol{\omega}_{1,\ldots,k}\right),$$

for all $i\in\mathcal{H}$ *and* $k=1,2,\ldots,n-1$, *where* $g(\boldsymbol{\omega}_{k+1,\ldots,n}|\boldsymbol{\omega}_{1,\ldots,k})$ *is the conditional prior belief on partial state* $\boldsymbol{\omega}_{k+1,\ldots,n}$ *given* $\boldsymbol{\omega}_{1,\ldots,k}$. *For the sake of completeness, we denote* $g(\boldsymbol{\omega}_{1,\ldots,k})=1$ *and* $g(\boldsymbol{\omega}_{k+1,\ldots,n}|\boldsymbol{\omega}_{1,\ldots,k})=g(\boldsymbol{\omega})$ *for* $k=0$.

The proof is presented in Appendix B.

It is noteworthy that $-\lambda_n\ln\Theta(p(i|\boldsymbol{\omega}))$ can be regarded as a correction term in the closed-form expression (24). Such a correction term is constructed endogenously according to the RI choice behaviour conditioned on the partial states. Whereas the correction terms in the exiting logit models, such as the link size logit model (Fosgerau et al., 2013), the nested recursive logit model (Mai et al., 2015), the path size logit model (Duncan et al., 2020), and the choice aversion logit model (Knise et al., 2022), and so on, are constructed using the exogenous ingredients that are independent of the choice behaviour. This distinct construction of correction term makes RI route choice model (21)-(23) differ from the logit models and its generalized variants in terms of prediction results.

Theorem 2 highlights the impact of the route travel costs, the marginal information costs, as well as the partial-information-based conditional route choice probabilities on the optimal RI route choice behaviour. In fact, $\Theta(p(i|\boldsymbol{\omega}))$ in Condition (24) is a multiplication of the rescaled conditional route choice probabilities based on partial information, where the exponent $(\lambda_k-\lambda_{k-1})/\lambda_n$ highlights the impact of the non-uniform marginal information costs on the RI route choice behaviour. As we can observe, $\Theta(p(i|\boldsymbol{\omega}))$ reverts to $p(i)$ when $\lambda_0=0$ and $\lambda_1=\lambda_2=\cdots=\lambda_n$. Thus, $\{\Theta(p(i|\boldsymbol{\omega}))\}$ is an extension of the habitual choice behaviour $\{p(i)\}$ in Condition (13).

Theorem 1 plays an indispensable role in characterizing the closed-form expression (24) by applying the KKT conditions to the route choice model (21)-(23). A vital step to establish the KKT conditions is to take partial derivatives on the Lagrangian of route choice model (21)-(23) with respect to each interior point $p(i|\boldsymbol{\omega})$. The Lagrangian incorporates the objective function (21) and the constraints (22)-(23) into a single equation, which quantifies the trade-off between the change of the total cost and the gain of updating the RI route choice behaviour. Theorem 1 ensures all the optimal conditional route choice probabilities locate within the interior of the feasible region. Thus, the KKT conditions guarantee that the optimal route choice probabilities can be achieved, only if the marginal trade-offs are zero with respect to all the $p(i|\boldsymbol{\omega})$'s, which gives Condition (24). The necessity is thus guaranteed. Conversely, the condition that $p(i|\boldsymbol{\omega})$ is positive enables us to choose zero Lagrangian multiplier for Constraint (23) so as to satisfy the complementary slackness condition. In addition, substituting Condition (24) into the Lagrangian of route choice model (21)-(23) make us possible to choose the proper Lagrangian multiplier for Constraint (22), such that the KKT conditions for route choice model (21)-(23) are satisfied. This guarantees the sufficiency. Thus, Condition (24) can be served as the closed-form expression of the optimal conditional route choice probabilities.

3.3 Solution Algorithm

Based on the closed-form expression (24), we devised a solution algorithm for route choice model (21)-(23).

Algorithm 1 Optimal Conditional Choice Probabilities

Step 1: Given prior beliefs $g(\cdot)$ on partial and full states and the marginal information costs $0<\lambda_0<\lambda_1\leq\lambda_2\leq\cdots\lambda_n<+\infty$. Start with initial conditional choice probabilities $\boldsymbol{p}^0=\{p^0(i|\boldsymbol{\omega})\}$ that satisfy $\boldsymbol{p}^0>0$. Set an iteration counter $t=0$ and an accuracy gap $\varepsilon>0$.

Step 2: According to $\boldsymbol{p}^t$, compute the unconditional route choice probability $p^t(i)$ for all $i\in\mathcal{H}$, and the route choice probability $p^t(i|\boldsymbol{\omega}_{1,\ldots,k})$ conditioned on partial state $\boldsymbol{\omega}_{1,\ldots,k}$ for all $i\in\mathcal{H}$ and $k=1,2,\ldots,n-1$. Then, update $\boldsymbol{p}^{t+1}$ as

$$p^{t+1}(i|\boldsymbol{\omega})=\frac{e^{-\frac{c(i,\boldsymbol{\omega})}{\lambda_n}}\Theta(p^t(i|\boldsymbol{\omega}))}{\sum_{j\in\mathcal{H}}e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda_n}}\Theta(p^t(j|\boldsymbol{\omega}))}. \tag{25}$$

Step 3: Check whether $\boldsymbol{p}^{t+1}$ satisfies

$$|p^{t+1}(i|\boldsymbol{\omega})-p^t(i|\boldsymbol{\omega})|<\varepsilon, \tag{26}$$

for all $i\in\mathcal{H}$ and $\boldsymbol{\omega}\in\boldsymbol{\Omega}$.

If yes, abort with $\boldsymbol{p}^{t+1}$ as the optimal solution.

Otherwise, go back to **Step 2**.

Note that the updating process in (25) guarantees $\{p(i|\boldsymbol{\omega})\}$ satisfies $\sum_{i\in\mathcal{H}}p^{t+1}(i|\boldsymbol{\omega})=1$ and $p^{t+1}(i|\boldsymbol{\omega})>0$ for all $i\in\mathcal{H}$ and $\boldsymbol{\omega}\in\Omega$. Criterion (26) is a relaxation of criterion $p^{t+1}(i|\boldsymbol{\omega})=p^t(i|\boldsymbol{\omega})$ so as to accelerate the convergence of Algorithm 1. In the following, we will prove that iteration (25) in Step 2 always reduces the objective $C_{\text{non}}(\boldsymbol{p})$ until $p^{t+1}(i|\boldsymbol{\omega})=p^t(i|\boldsymbol{\omega})$ is satisfied for all $i\in\mathcal{H}$ and $\boldsymbol{\omega}\in\Omega$.

Theorem 3. *Under Assumption 1, Algorithm 1 converges to the optimal solution of route choice model (21)-(23) when* $p^{t+1}(i|\boldsymbol{\omega}) = p^t(i|\boldsymbol{\omega})$ *for all* $i \in \mathcal{H}$ and $\boldsymbol{\omega} \in \boldsymbol{\Omega}$.

The proof be derived simply by combining Theorem 2 and Proposition 2 in Huettner et al. (2019).

Theorem 3 ensures that Algorithm 1 converges to the optimal solutions of route choice model (21)-(23). In the existing works of Caplin et al. (2019), Huettner et al. (2019), Jiang et al. (2020) and Walker-Jones (2023), the structure of optimal conditional choice probabilities can only be served as necessary but insufficient condition for optimality. This gives rise to the consequence that a series of additional necessary and sufficient conditions must be introduced to guarantee the optimality of the solutions of their choice models, e.g., Proposition 1 in Caplin et al. (2019), Theorem 2 in Huettner et al. (2019), Proposition 2 in Jiang et al. (2020) and Theorem 3 in Walker-Jones (2023). In this study, because of the characterization of the closed-form expression (24), a much simpler criterion (26) is enough to check the convergence of Algorithm 1.

In terms of the behavioural implications, the RI model provides a framework to model how information can influence traveller's route choice and lead them towards more optimal choices. More specifically, when the information on the states are given, the RI traveller would formulate the initial RI route choice behaviour according to his individual cognition, which may not be optimal. The solution algorithm provides a feasible way for the RI traveller to learn the optimal route choice behaviour from the initial one. In Algorithm 1, the set of states and the initial route choice behaviour (the initial conditional choice probabilities) are considered as inputs. Then, the RI traveller keeps adjusting the conditional route choice probabilities according to the rule of updating, until the optimality criterion is satisfied for all routes and states. From which, the optimal route choice behaviour can be identified by the RI traveller.

4 Numerical examples

In this section, we present two numerical examples to illustrate the theoretical findings presented in the previous section. The first example is conducted on a toy network, which is designed to illustrate the convergence performance of the solution algorithm and to highlight the impact of marginal information costs on the optimal conditional route choice probabilities. The second example is conducted in the Nguyen-Dupuis network, which demonstrates the effectiveness of the theoretical results over a large-scale route choice model. Algorithm 1 is coded in Matlab R2023a on a Windows machine with Intel(R) Core(TM) i9-11950H and 64 GB RAM. The initial conditional choice probabilities $\boldsymbol{p}^0$ in Algorithm 1 are randomly selected according to Constraints (22) and (23).

4.1 A toy network

Consider a toy network shown in Figure 1, which contains two candidate routes. Thus, the set of candidate routes is defined by $\mathcal{H} = \{1, 2\}$.

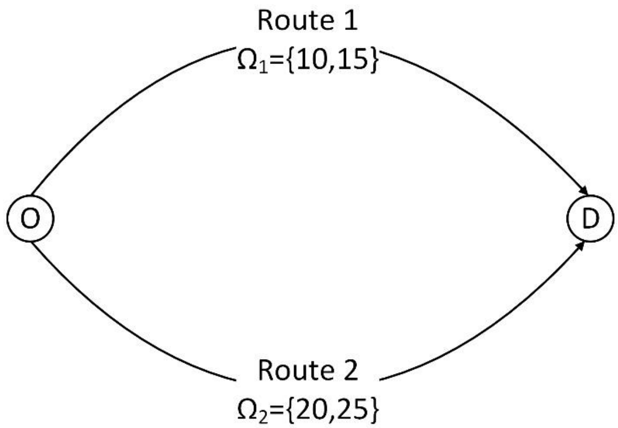


Figure 1: A toy network for illustration.

We assume that there are two (fixed) travel costs for each route associated with the states of traffic conditions: non-congested and congested. The traffic conditions are noted in Figure 1, where the first value in the bracket for each route is the travel cost in the non-congested state, whilst the second value is the travel cost in the congested state. The set of sub-components are denoted by $\Omega_1 = \{10, 15\}$ and $\Omega_2 = \{20, 25\}$ respectively. Then, the set of all possible states is denoted by $\boldsymbol{\Omega} = \{(10, 20), (15, 20), (10, 25), (15, 25)\}$. For presentation convenience, we denote the states in $\boldsymbol{\Omega}$ by $\boldsymbol{\omega}^1 = (10, 20)$, $\boldsymbol{\omega}^2 = (15, 20)$, $\boldsymbol{\omega}^3 = (10, 25)$, $\boldsymbol{\omega}^4 = (15, 25)$. The corresponding travel costs in these states are given by $c(1, \boldsymbol{\omega}^j) = \xi_1^j$, $c(2, \boldsymbol{\omega}^j) = \xi_2^j$, where ξ_1^j and ξ_2^j are the first and second sub-component of state $\boldsymbol{\omega}^j$ respectively. We assume the two sub-components in each state occur independently with equal probability. The accuracy gap ε is set to be 10^{-10}, that is, Algorithm 1 terminates if $|p^{t+1}(i|\boldsymbol{\omega}) - p^t(i|\boldsymbol{\omega})| < 10^{-10}$ is satisfied for all $i \in \mathcal{H}$ and $\boldsymbol{\omega} \in \boldsymbol{\Omega}$.

First, we test the impact of λ_0 on the optimal conditional route choice probabilities. To this end, we fix the values of λ_1 and λ_2 at 15 and 21 respectively, and take the values of λ_0 from the set $\{5, 7, 9, 11\}$. The simulation results are presented in Figures 2 and 3. Figure 2(a)-(d) depict the convergence performance of the conditional route choice probabilities for Route 1 in each specified state for the different values of λ_0. The convergence performance of the conditional route choice probabilities for Route 2 are omitted, because they can be simply calculated by $p(2|\boldsymbol{\omega}) = 1 - p(1|\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \boldsymbol{\Omega}$. Note that the optimal conditional route choice probabilities with $\lambda_0 = 0$ is computed by Algorithm 1 in Huettner et al. (2019). Figure 3 depicts the amount of acquired information from Ω_0 for the different values of λ_0.

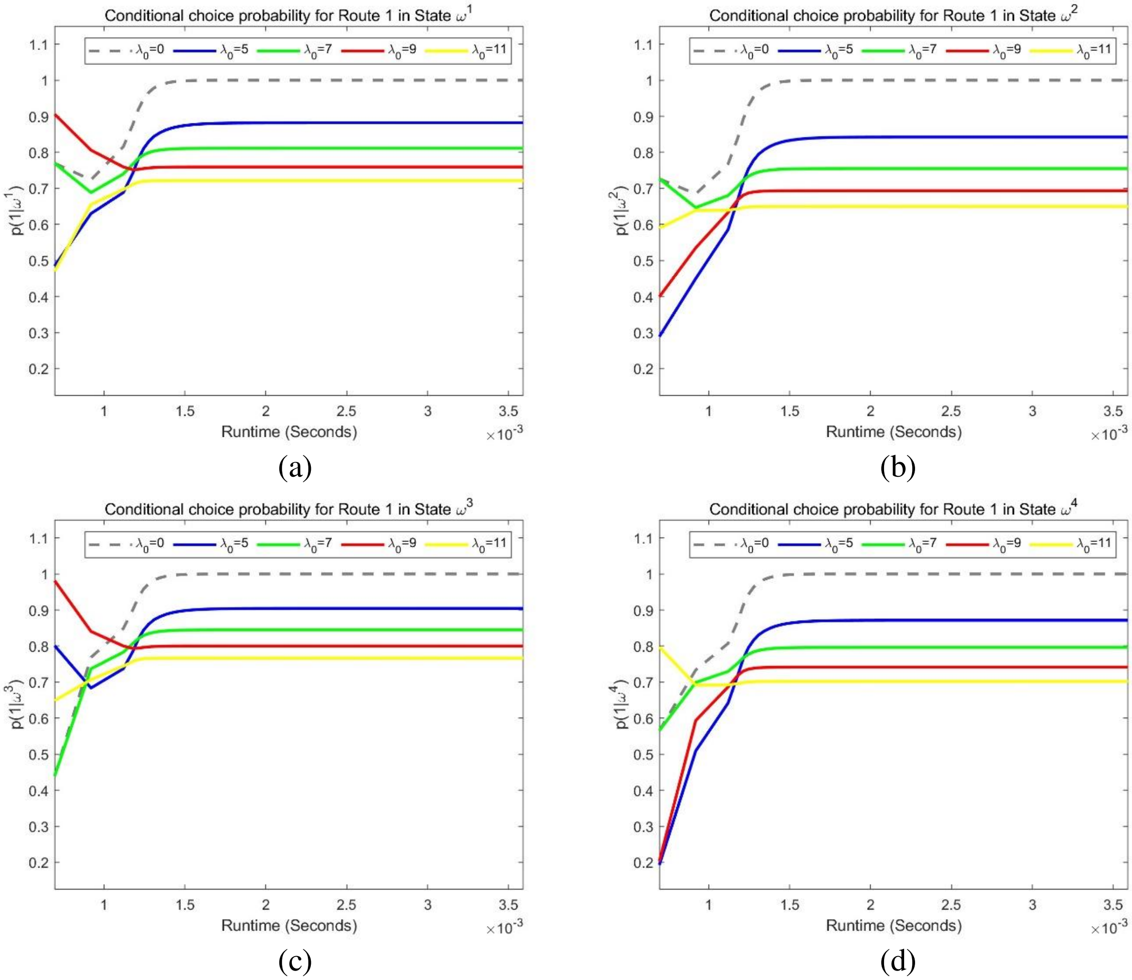


Figure 2: Convergence performance of conditional route choice probabilities for Route 1 in each specific state with $\lambda_0 = 5, 7, 9, 11$: (a) $p(1|\boldsymbol{\omega}^1)$, (b) $p(1|\boldsymbol{\omega}^2)$, (c) $p(1|\boldsymbol{\omega}^3)$, (d) $p(1|\boldsymbol{\omega}^4)$.

We observe from Figure 2 that the conditional route choice probabilities for Route 1 converge to the optimal values in all states, which verifies the convergence of Algorithm 1. In addition, it is noteworthy that the travel cost for Route 2 is higher than that for Route 1 in all states, which means that Route 2 is dominated by Route 1. When $\lambda_0 = 0$, Figure 2 shows that the optimal conditional route choice probabilities for Route 1 are always one in all states, which are consistent with the results in Huettner et al. (2019), Jiang et al. (2020), Matějka and Mckay (2015) and Walker-Jones (2023) that all the optimal conditional route choice probabilities locate on the boundary of the feasible region. However, when λ_0 is positive, we can observe from Figure 2 that the optimal conditional route choice probabilities for Route 1 are strictly less than one in all states, which means that all the optimal conditional route choice probabilities locate within the interior of the feasible region. This comparison demonstrates that introducing the background information before launching the information acquisition process gives rise to a substantial change of the RI route choice behaviour.

From an alternative perspective, we observe from Figure 2(a) that the optimal value of $p(1|\boldsymbol{\omega}^1)$ decreases, and correspondingly, the optimal value of $p(2|\boldsymbol{\omega}^1)$ increases, as λ_0 increases. Simultaneously, the same changing trends for the optimal values of $p(1|\boldsymbol{\omega}^j)$ for

$j = 2, 3, 4$ can also be observed in Figures 2(b)-2(d). Moreover, as illustrated in Figure 3, the amount of acquired information from Ω_0 decreases markedly with respect to increasing λ_0 for fixed $\lambda_1 = 15$ and $\lambda_2 = 21$. Combining the observations from Figure 2 and 3, we find that larger amount of acquired information from Ω_0 gives rise to higher conditional route choice probabilities for Route 1, and vice versa. The reason is that when more information from Ω_0 is acquired, the RI traveller would be more confident that Route 1 is with lower travel cost than Route 2, and correspondingly, choose Route 1 with a higher chance.

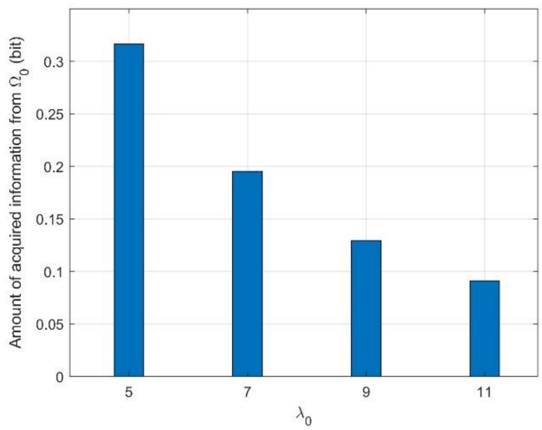
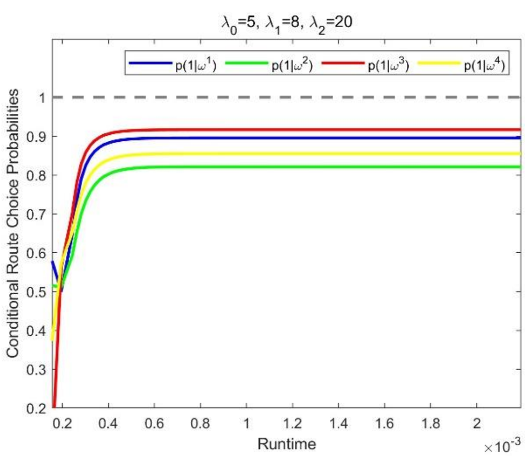
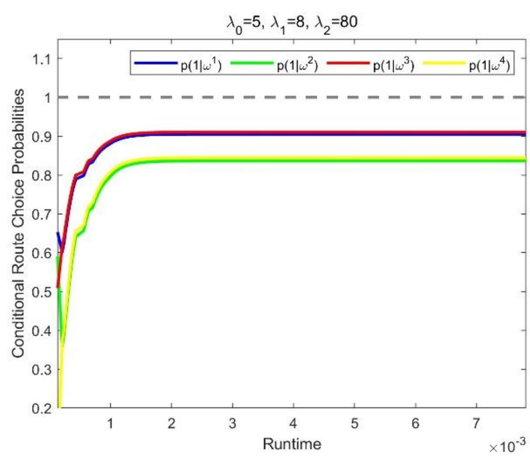


Figure 3: Amount of acquired information from Ω_0 with respect to different values of λ_0.

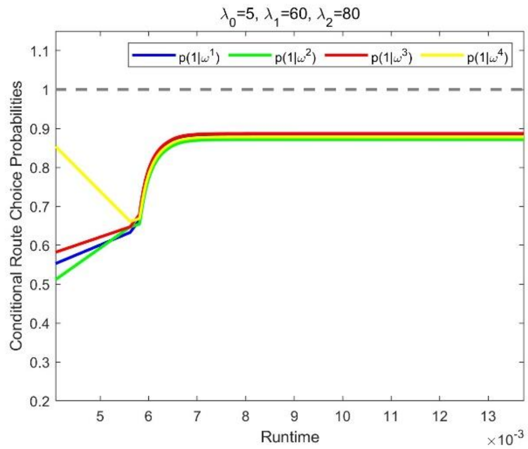
Next, we test the impact of λ_1 and λ_2 on the optimal conditional route choice probabilities. To this end, we fix the value of λ_0 at 5, and take the values of (λ_1, λ_2) from the set $\{(8, 20), (8, 80), (60, 80)\}$. The simulation results are presented in Figure 4 and 5. Figure 4(a)-(c) depict the convergence performance of the conditional route choice probabilities for Route 1 in all states for each specified λ_1 and λ_2. Figure 5 depicts the amount of acquired information from Ω_1 and Ω_2 for the different values of λ_1 and λ_2.



(a)



(b)



(c)

Figure 4: Convergence performance of the conditional route choice probabilities for Route 1 in the four states for specified marginal information costs. (a) $\lambda_0 = 5$, $\lambda_1 = 8$, $\lambda_2 = 20$, (b) $\lambda_0 = 5$, $\lambda_1 = 8$, $\lambda_2 = 80$, (c) $\lambda_0 = 5$, $\lambda_1 = 60$, $\lambda_2 = 80$.

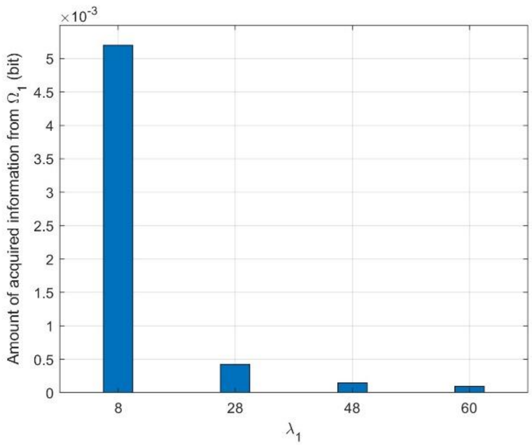
When $\lambda_0 = 5$, $\lambda_1 = 8$ and $\lambda_2 = 20$, we can observe from Figure 4(a) that the conditional route choice probabilities for Route 1 exhibit significantly discrepancies in different states, i.e. $p(1|\boldsymbol{\omega}^1) > p(1|\boldsymbol{\omega}^3) > p(1|\boldsymbol{\omega}^2) > p(1|\boldsymbol{\omega}^4)$, because the mild marginal information costs enable the RI traveller acquire adequate information to clearly distinguish the travel costs for both routes, i.e. the information acquired from Ω_0 , Ω_1 and Ω_2 are good enough.

When $\lambda_0 = 5$, $\lambda_1 = 8$ and $\lambda_2 = 80$, a different result can be observed from Figure 4(b), which illustrates that $p(1|\boldsymbol{\omega}^1) \approx p(1|\boldsymbol{\omega}^3)$ and $p(1|\boldsymbol{\omega}^2) \approx p(1|\boldsymbol{\omega}^4)$. This observation indicates that the travel costs for Route 2 can not be identified, which means that the information acquired from Ω_0 and Ω_1 are good enough and the effort in acquiring information from Ω_2 is not worth it. This statement can be demonstrated by Figure 5(b), which illustrates the amount of acquired information from Ω_2 with respect to four different values of λ_2 for fixed $\lambda_0 = 5$ and $\lambda_1 = 8$. We can observe that the amount of acquired information from Ω_2 decreases as λ_2 increases and the amount of acquired information from Ω_2 is only 5.32×10^{-5} bit when $\lambda_2 = 80$. The comparison between Figure 4(a) and Figure 4(b) demonstrates that there exists a cutoff point of λ_2, which can be observed by gradually increasing it from 20 to 80 .

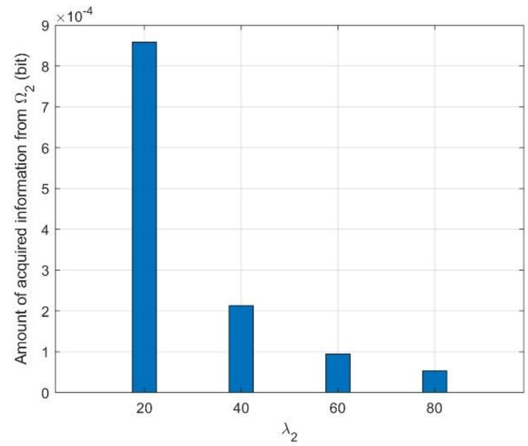
When $\lambda_0 = 5$, $\lambda_1 = 60$ and $\lambda_2 = 80$, the observations from Figure 4(c) yields a consequence that $p(1|\boldsymbol{\omega}^1) \approx p(1|\boldsymbol{\omega}^2) \approx p(1|\boldsymbol{\omega}^3) \approx p(1|\boldsymbol{\omega}^4)$, which indicate that the RI traveller can only identify the habitual information. This observation means that the information acquired from Ω_0 is good enough and the effort in acquiring information from Ω_1 and Ω_2 is not worth it. This statement can be demonstrated by Figure 5(a), which illustrates the amount of acquired information from Ω_1 with respect to four different values of λ_1 for fixed $\lambda_0 = 5$ and $\lambda_2 = 80$. We can observe that the amount of acquired information decreases as λ_1 increases and the amount of acquired information from Ω_1 is only 9.2×10^{-5} bit when $\lambda_1 = 60$. The comparison between

Figure 4(b) and Figure 4(c) demonstrates that there exists a cutoff point of λ_1, which can be observed by gradually increasing it from 8 to 60.

From the observations in Figure 2 to Figure 5, we find that the marginal information costs impact the RI route choice behaviour through controlling the amount of acquired information from the corresponding set of sub-components.



(a)



(b)

Figure 5: The amount of acquired information from each set of sub-components. (a) The first sub-component Ω_1, (b) The second sub-component Ω_2.

Then, we consider the RI route choice behaviour with $\lambda_0 = 400$, $\lambda_1 = 500$ and $\lambda_2 = 600$, which represents an extreme case that the marginal information costs are too high for the RI traveller to afford. As a result, both habitual information and the sub-component information become too expensive for the RI traveller to acquire. Figure 6 illustrates the numerical result, in which the optimal conditional route choice probabilities are nearly $1/2$ in all states, whilst the total amount of acquired information is 1.34×10^{-4} bit. This observation demonstrates that if no information is available, the RI traveller would choose the candidate routes with equal probability in all states. This result is consistent with the principle of maximum entropy introduced in Section 3.1.

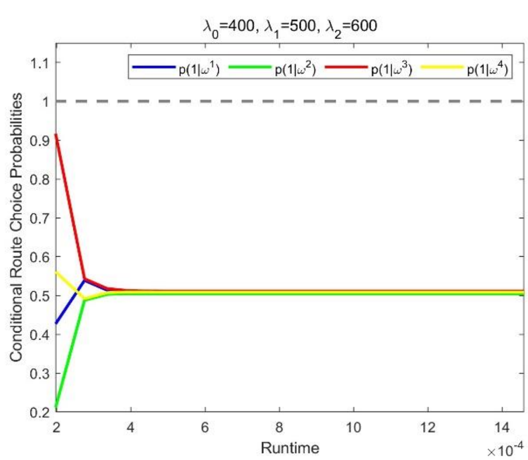
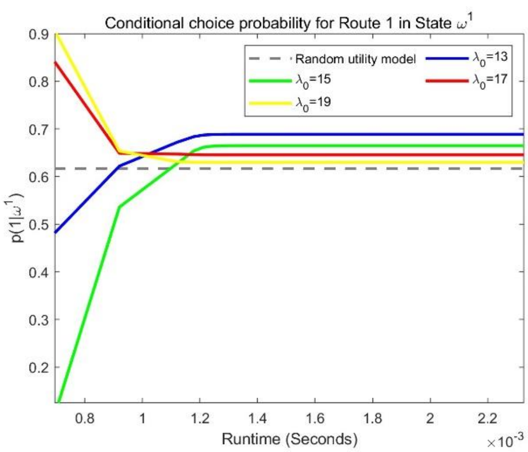
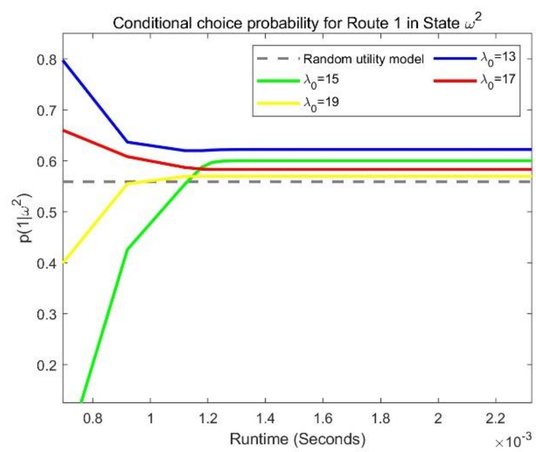


Figure 6: Convergence performance of the conditional route choice probabilities for Route 1 in the four states with $\lambda_0 = 400$, $\lambda_1 = 500$, $\lambda_2 = 600$.

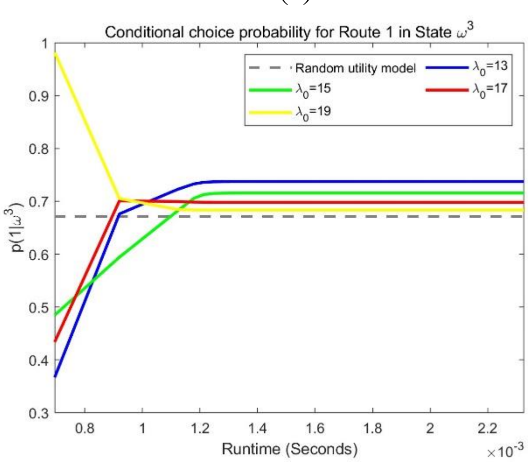
Figure 7 illustrate the comparisons of the optimal conditional route choice probabilities between RI route choice model (21)-(23) and the corresponding random utility choice model. We compute the optimal conditional route choice probabilities of RI route choice model (21)-(23) for Route 1 with $\lambda_0 = 13, 15, 17, 19$, $\lambda_2 = 21$ and $\lambda_1 = (\lambda_0 + \lambda_2)/2$. We also compute the optimal choice probability of the random utility model for Route 1 in each specific state, in which the random disturbances for the route travel costs are assumed to follow the i.i.d. Gumbel distribution with scale parameter $\lambda = 21$. In Figure 7, the grey dash lines depict the route choice behaviour described by the random utility model, whilst the colour lines depict the RI route choice behaviours. We can observe from Figure 7 that RI route choice behaviour gradually approaches to the random-utility-based route choice behaviour as λ_0 approaches to λ_2. This observation infers that RI route choice model (21)-(23) would describe a random-utility-based route choice behaviour when $\lambda_0 = \lambda_1 = \lambda_2$. In fact, by letting $\lambda_0 = \lambda_1 = \cdots = \lambda_n$, closed-form expression (24) degenerates into $p(i|\boldsymbol{\omega}) = e^{-\frac{c(i,\boldsymbol{\omega})}{\lambda_n}} / \sum_{j \in \mathcal{H}} e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda_n}}$, which is consistent with the optimal choice probability of a random utility model. This result not only validates the above inference on the observations in Figure 7, but also demonstrates that the RI choice behaviour characterized by closed-form expression (24) always contains one random-utility-based choice behaviour as a special case.



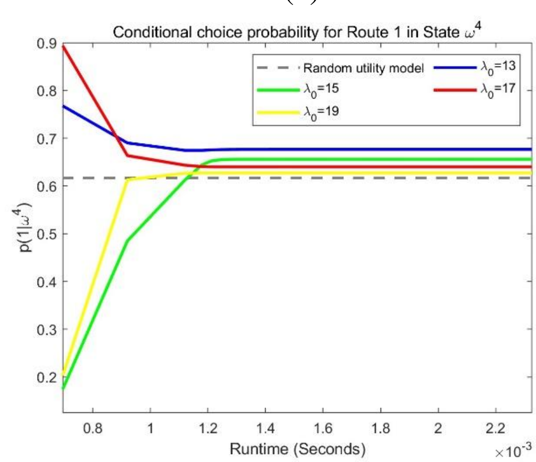
(a)



(b)



(c)



(d)

Figure 7: Comparisons of the optimal conditional route choice probabilities between RI route choice model (21)-(23) and the corresponding random utility model: (a) $p(1|\boldsymbol{\omega}^1)$, (b) $p(1|\boldsymbol{\omega}^2)$, (c) $p(1|\boldsymbol{\omega}^3)$, (d) $p(1|\boldsymbol{\omega}^4)$.

4.2 Nguyen-Dupius Network

In this subsection, we adopt the Nguyen-Dupius (ND) network to further illustrate the RI choice behaviour described by the route choice model (21)-(23). Figure 8 depicts the layout of the ND network, where the circles represent the nodes and the solid lines with arrows represent the directed links. In total, the ND network contains 4 O-D pairs, 19 links and 25 routes. The links are labelled by the numbers above them. Table 1 illustrates the link-route relationship of the ND network. Similar to Section 3.1, we assume that there are two (fixed) travel costs for each link associated with the states, which are shown in the brackets under each link in Figure 8. We further assume that these states as well as their sub-components occur independently with equal probability. There are therefore totally 2^{19}=524288 possible states, each of them contains 19 sub-components. The route travel cost in each state can then be calculated as the accumulation of the link cost in the state according to the link-route relationship.

We consider four scenarios of marginal information costs. In each scenario, $\lambda_0, \lambda_1, ..., \lambda_{19}$ are assumed to be an arithmetic sequence. For scenario j, we set $\lambda_0 = j$, $\lambda_{19} = 20j$ and common difference be j. Since this problem encounters a large amounts of states, we set the accuracy gap be 10^{-3}, that is, Algorithm 1 terminates if $|p^{t+1}(i|\boldsymbol{\omega}) - p^t(i|\boldsymbol{\omega})| < 10^{-3}$ is satisfied for all $i \in \mathcal{H}$ and $\boldsymbol{\omega} \in \boldsymbol{\Omega}$.

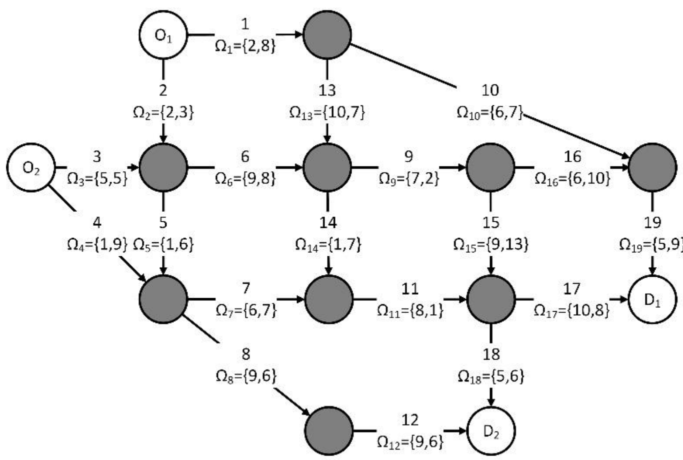
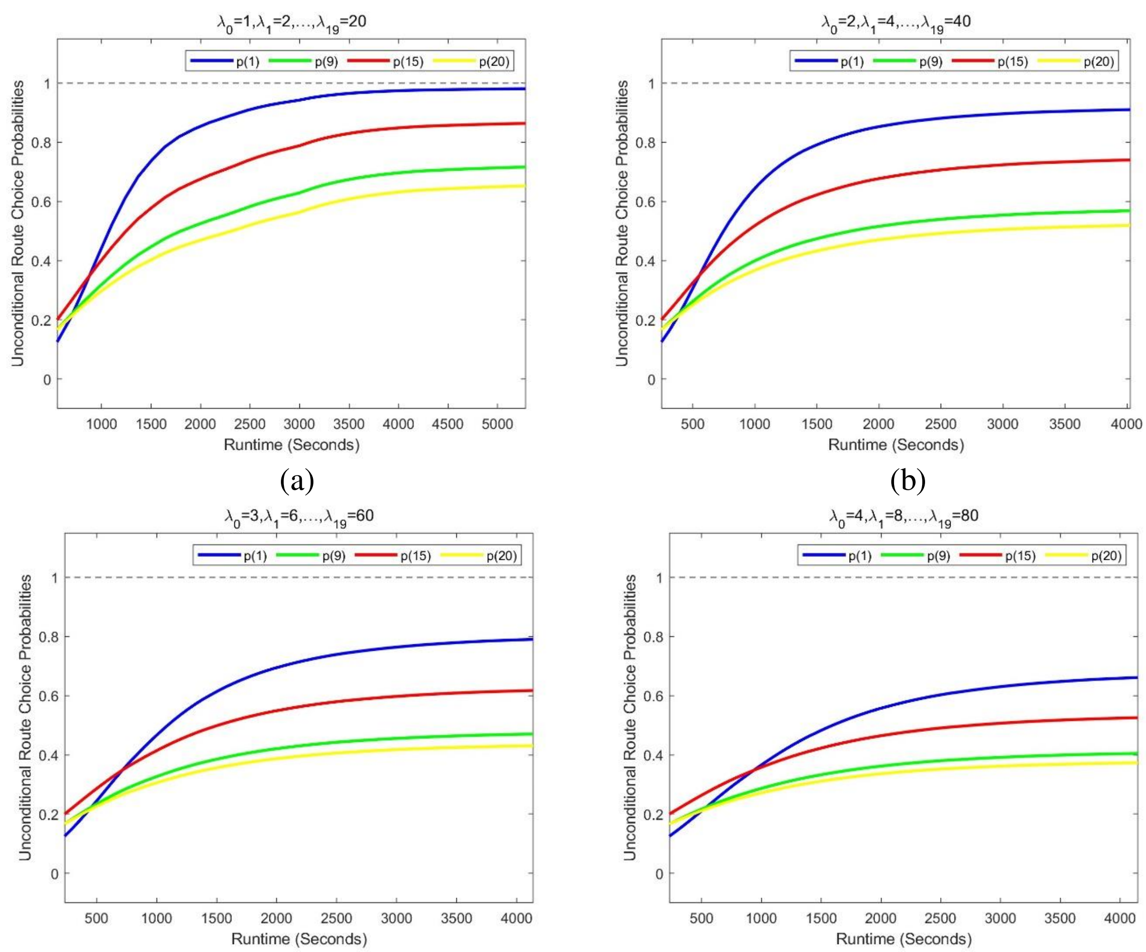


Figure 8: The Nguyen-Dupius Network.

Table 1: Link-Route incidence relationship of the ND network.

O-D pair	Route no.	Link sequence	O-D pair	Route no.	Link sequence
(O_1, D_1)	1	1, 10, 19	(O_1, D_2)	9	2, 5, 8, 12
	2	2, 6, 9, 16, 19		10	2, 6, 9, 15, 18
	3	2, 6, 9, 15, 17		11	2, 6, 14, 11, 18
	4	2, 6, 14, 11, 17		12	2, 5, 7, 11, 18

	5	2, 5, 7, 11, 17		13	1, 13, 9, 15, 18
	6	1, 13, 9, 16, 19		14	1, 13, 14, 11, 18
	7	1, 13, 9, 15, 17		20	4, 8, 12
	8	1, 13, 14, 11, 17		21	4, 7, 11, 18
(O_2, D_1)	15	4, 7, 11, 17	(O_2, D_2)	22	3, 5, 8, 12
	16	3, 6, 14, 11, 17		23	3, 6, 9, 15, 18
	17	3, 6, 9, 16, 19		24	3, 6, 14, 11, 18
	18	3, 6, 9, 15, 17		25	3, 5, 7, 11, 18
	19	3, 5, 7, 11, 17			



(a) (b) (c) (d)

Figure 9: Optimal unconditional route choice probabilities for some typical routes. (a). Route 1 between (O_1, D_1); (b). Route 9 between (O_1, D_2); (c). Route 15 between (O_2, D_1); (d). Route 20 between (O_2, D_2).

Figure 9 depicts the unconditional route choice probabilities for Routes 1, 9, 15 and 20. These routes are with the lowest travel costs between the corresponding O-D pairs in all states, and thus their unconditional route choice probabilities are the highest. We observe from Figure 9 that the convergence of Algorithm 1 is reached and within a reasonable time (4000 or 5000 seconds) for a network with a large number (2^{19}) of decision variables. In addition, we observe that the optimal values of these unconditional route choice probabilities for specific route decrease as the marginal information costs increase. For instance, by comparing Figures 9(a) and 9(b), we can see that the optimal values of $p(1)$ in Scenario 1 is greater than which in Scenario 2. Similar observations can be obtained by comparing the optimal values of $p(1)$ in Figures 9(b) and 9(c) as well as in Figures 9(c) and 9(d). Correspondingly, it can be observed in Figure 10 that the total amount of acquired information decreases as marginal information costs increase. These observations are consistent with the consequence exemplified through a toy network in Section 4.1, which states that high marginal information costs make the RI traveller acquires less information, and thus, more difficult to distinguish the different travel costs for these routes in each state. Therefore, we demonstrate that the marginal information costs can also impact the RI route choice behaviour in a large-scale route choice model to a great extent.

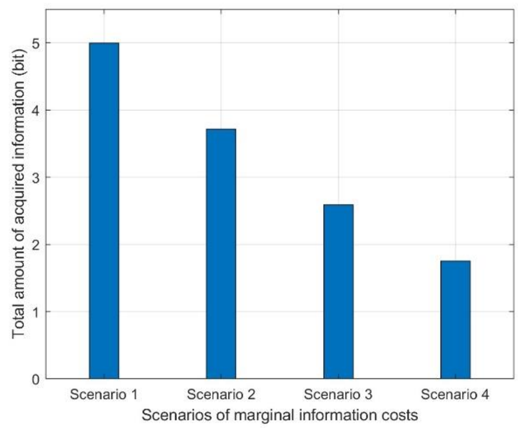


Figure 10: Total amount of acquired information for four different scenarios of marginal information costs.

5 Conclusions and future research directions

This study focuses on the RI route choice behaviour by taking the non-uniform marginal information costs into account. Unlike the traditional choice models based on the random utility model framework, which rely on exogenous random shocks, our model considers the endogenous incentive for the RI traveller to acquire information on the states throughout the transportation network. This novel perspective sheds light on the impact of marginal information costs on the RI route choice behaviour. We establish a generalized RI route choice model with non-uniform marginal information costs, which incorporates the background information that contains all the candidate routes. These candidate routes are flexibly given, which can be considered either as the whole feasible routes of the entire network, or the routes that the traveller knows about.

We prove that such generalized RI route choice model ensures that all the optimal conditional route choice probabilities locate within the interior of the feasible region, that is, all the candidate route are assigned a positive choice probability in all states. This behaviour differs significantly to the RI route choice behaviour described in the existing works, the latter allows a set of candidate routes to remain unchosen. This property enables us to characterize the closed-form expression of the optimal conditional choice probabilities and devise an iterative solution algorithm. Whereas there is no closed-form expression of the optimal conditional choice probabilities in the existing works. Our work also presents a feasible way to characterize the closed-form expression of the choice behaviour in the general RI choice modelling framework.

It is worth noting that, although our work considers the route choice behaviour from the perspectives of individual RI traveller, this RI route choice modelling framework can also be applicable from the perspective of the system operator (including that for the AVs and navigation apps). In the era of AVs and navigation apps, travellers may no longer need to acquire information by themselves. Instead, the information acquisition can be done by the navigation system. Such system-led process is still costly, that is, the navigation system has to consume its limited resources, such as storage, computing power and communication, to acquire information on the traffic conditions and to generate recommendation (i.e., optimal route choice probability) for its users. Our proposed RI framework can also be used by the navigation system to formulate the recommendation of travelling routes.

The proposed model has some limitations that warrant further investigation. Firstly, the model does not explicitly consider the timing of information acquisition process. In reality, the RI traveller needs not only determine what and how much information to acquire, but also when to do so. Thus, incorporating a dynamic route choice model that accounts for the timeliness of information acquisition would provide a more realistic representation of RI route choice behaviour in this information-rich society. Secondly, the solution algorithm (Algorithm 1) is computationally challenging when facing large-scale RI route choice models. In the ND network, it takes more than 4000 seconds to solve the RI route choice model of 524288 possible states. Addressing this computational challenge will help pave the way for practical application of the RI route choice model. In addition, the marginal information costs in the choice model require further calibration to align with different engineering application scenarios. While it may be intuitive to estimate these costs from historical data using statistical estimation methods such as maximum likelihood estimation, the calibration of marginal information costs has not been thoroughly explored in existing RI models. Future research should focus on developing methodologies to accurately estimate these costs and tailor them to specific contexts. Addressing these limitations will enhance the practical applicability and realism of the proposed model, allowing for a better understanding of the dynamics of RI route choice behaviour and facilitating its implementation in real-world transportation networks.

CRediT authorship contribution statement

Bo Zhou: Conceptualization, Methodology, Funding acquisition, Validation, Formal analysis, Investigation, Writing-review & editing. **Ronghui Liu**: Conceptualization, Methodology, Validation, Supervision, Formal analysis, Writing-review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank the reviewers and the editor for their valuable suggestions and comments which have led to a much improved paper. This work is supported in part by the National Natural Science Foundation of China [Grant 61803056].

Data availability

No data was used for the research described in the article.

Appendix A.

Proof of Theorem 1. By contradiction, we suppose that there exist optimal solutions $\boldsymbol{p}^* = \{p^*(i|\boldsymbol{\omega})\}$ of route choice model (21)-(23) such that $p^*\left(\bar{i}|\breve{\boldsymbol{\omega}}\right) = 0$ for certain $\bar{i} \in \mathcal{H}$ and $\breve{\boldsymbol{\omega}} \in \Omega$. By Constraint (22), there must exist a route $\hat{i} \in \mathcal{H}$ other than $\bar{i}$ such that $p^*\left(\hat{i}|\breve{\boldsymbol{\omega}}\right) > 0$. In the following, we construct another feasible solution $\{\tilde{p}(i|\boldsymbol{\omega})\}$ by perturbing the conditional route choice probabilities $p^*\left(\bar{i}|\breve{\boldsymbol{\omega}}\right)$ and $p^*\left(\hat{i}|\breve{\boldsymbol{\omega}}\right)$, whilst other conditional route choice probabilities remain unchanged. To this end, we define $\tilde{\boldsymbol{p}} = \{\tilde{p}(i|\boldsymbol{\omega})\}$ as follows

$$\tilde{p}\left(\bar{i}|\breve{\boldsymbol{\omega}}\right) = p^*\left(\bar{i}|\breve{\boldsymbol{\omega}}\right) + \epsilon,$$

$$\tilde{p}\left(\hat{i}|\breve{\boldsymbol{\omega}}\right) = p^*\left(\hat{i}|\breve{\boldsymbol{\omega}}\right) - \epsilon,$$

$$\tilde{p}\left(\bar{i}|\boldsymbol{\omega}\right) = p^*\left(\bar{i}|\boldsymbol{\omega}\right), \text{ for all } \boldsymbol{\omega} \in \Omega \setminus \{\breve{\boldsymbol{\omega}}\},$$

$$\tilde{p}\left(\hat{i}|\boldsymbol{\omega}\right) = p^*\left(\hat{i}|\boldsymbol{\omega}\right), \text{ for all } \boldsymbol{\omega} \in \Omega \setminus \{\breve{\boldsymbol{\omega}}\},$$

$$\tilde{p}(i|\boldsymbol{\omega}) = p^*(i|\boldsymbol{\omega}), \text{ for all } i \in \mathcal{H} \setminus \left\{\bar{i}, \hat{i}\right\}, \boldsymbol{\omega} \in \Omega,$$

where $\epsilon > 0$ is a sufficiently small perturbation.

Then, we make a difference between $C_{\text{non}}(\tilde{\boldsymbol{p}})$ and $C_{\text{non}}(\boldsymbol{p}^*)$, which is exhibited as follows

$$\begin{aligned} &C_{\text{non}}(\tilde{\boldsymbol{p}}) - C_{\text{non}}(\boldsymbol{p}^*) \\ &= \underbrace{\sum_{\boldsymbol{\omega} \in \Omega} c\left(\bar{i}, \boldsymbol{\omega}\right)\left(\tilde{p}\left(\bar{i}|\boldsymbol{\omega}\right) - p^*\left(\bar{i}|\boldsymbol{\omega}\right)\right) g(\boldsymbol{\omega}) + \sum_{\boldsymbol{\omega} \in \Omega} c\left(\hat{i}, \boldsymbol{\omega}\right)\left(\tilde{p}\left(\hat{i}|\boldsymbol{\omega}\right) - p^*\left(\hat{i}|\boldsymbol{\omega}\right)\right) g(\boldsymbol{\omega})}_{\Delta_1} \\ &\quad + \underbrace{C_{\text{non,info}}\left(\tilde{\boldsymbol{p}}; \bar{i}\right) - C_{\text{non,info}}\left(\boldsymbol{p}^*; \bar{i}\right)}_{\Delta_2} + \underbrace{C_{\text{non,info}}\left(\tilde{\boldsymbol{p}}; \hat{i}\right) - C_{\text{non,info}}\left(\boldsymbol{p}^*; \hat{i}\right)}_{\Delta_3}, \end{aligned}$$

where $C_{\text{non,info}}(\boldsymbol{p}; i) = \sum_{k=0}^{n} \lambda_k \left(\mathscr{H}(\boldsymbol{p}; i|\boldsymbol{\Omega}_{-1,\dots,k-1}) - \mathscr{H}(\boldsymbol{p}; i|\boldsymbol{\Omega}_{-1,\dots,k})\right)$

Then, following a similar way to Du et al. (2014), if $p^*\left(\bar{i}\right) > 0$, then we can derive upper bounds M_1, M_2 and M_3 for Δ_1, Δ_2 and Δ_3, respectively, such that

$$C_{\text{non}}(\tilde{\boldsymbol{p}}) - C_{\text{non}}(\boldsymbol{p}^*) < \epsilon g(\breve{\boldsymbol{\omega}})\left(\ln \epsilon^{\lambda_n} + M_1 + M_2 + M_3\right);$$

if $p^*(\bar{i}) = 0$, then we can derive upper bounds M_1, M_2' and M_3 for Δ_1, Δ_2 and Δ_3, respectively, such that

$$C_{\text{non}}(\tilde{\boldsymbol{p}}) - C_{\text{non}}(\boldsymbol{p}^*) < \epsilon g(\breve{\boldsymbol{\omega}})\left(\ln \epsilon^{\lambda_0} + M_1 + M_2' + M_3\right).$$

Note that the upper bounds M_1, M_2, M_2' and M_3 are independent of ϵ. Thus, we can properly choose an ϵ such that

$$C_{\text{non}}(\tilde{\boldsymbol{p}}) - C_{\text{non}}(\boldsymbol{p}^*) < 0,$$

which contradicts to $\{p^*(i|\boldsymbol{\omega})\}$ are optimal solutions of route choice model (21)-(23).

Appendix B.

Proof of Theorem 2. The necessity part of proof can be obtained according to the derivation of Theorem 1 in Huettner et al. (2019). Thus, we omit it. In the following, we focus on the sufficiency part of proof.

If $p(i|\boldsymbol{\omega}) > 0$ and satisfies (24) for all $i \in \mathcal{H}$ and $\boldsymbol{\omega} \in \Omega$, then by substituting (24) into $C_{\text{non}}(\boldsymbol{p})$, we can get that

$$\begin{aligned} C_{\text{non}}(\boldsymbol{p}) = & \sum_{i \in \mathcal{H}} \sum_{\boldsymbol{\omega} \in \Omega} c\,(i,\boldsymbol{\omega})\, p(i|\boldsymbol{\omega})\, g(\boldsymbol{\omega}) + \sum_{k=0}^{n-1} \lambda_k \mathscr{I}(\boldsymbol{p}; \Omega_k | \boldsymbol{\Omega}_{-1,\ldots,k-1}) \\ & + \lambda_n \mathscr{H}(\boldsymbol{p}|\boldsymbol{\Omega}_{-1,\ldots,n-1}) \\ & + \lambda_n \sum_{\boldsymbol{\omega} \in \Omega} g\,(\boldsymbol{\omega}) \sum_{i \in \mathcal{H}} p\,(i|\boldsymbol{\omega}) \log \frac{e^{-\frac{c(i,\boldsymbol{\omega})}{\lambda_n}} \Theta(p(j|\boldsymbol{\omega}))}{\sum_{j \in \mathcal{H}} e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda_n}} \Theta(p(j|\boldsymbol{\omega}))}. \end{aligned}$$

By logarithm calculation and cancellation, we get an alternative formulation of $C_{\text{non}}(\boldsymbol{p})$

$$C_{\text{non}}(\boldsymbol{p}) = -\lambda_n \sum_{\boldsymbol{\omega} \in \Omega} g\,(\boldsymbol{\omega}) \ln\left(\sum_{j \in \mathcal{H}} e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda_n}} \Theta(p(j|\boldsymbol{\omega}))\right) + \lambda_0 \ln|\mathcal{H}|.$$

Applying Theorem 2 in Huettner et al. (2019), we get $C_{\text{non}}(\boldsymbol{p})$ is convex. Then, taking partial derivative of $C_{\text{non}}(\boldsymbol{p})$ with respect to $p(i|\boldsymbol{\omega})$ yields that

$$\frac{\partial C_{\text{non}}(\boldsymbol{p})}{\partial p(i|\boldsymbol{\omega})} = -\lambda_n g(\boldsymbol{\omega}) \ln\left(\sum_{j \in \mathcal{H}} e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda_n}} \Theta(p(j|\boldsymbol{\omega}))\right) + \lambda_0.$$

Let the Lagrangian multiplier for Constraint (22) be

$$\xi_1(\boldsymbol{\omega}) = \lambda_n g(\boldsymbol{\omega}) \ln\left(\sum_{j \in \mathcal{H}} e^{-\frac{c(j,\boldsymbol{\omega})}{\lambda_n}} \Theta(p(j|\boldsymbol{\omega}))\right) - \lambda_0,$$

and let the Lagrangian multiplier for Constraint (23) be

$$\xi_2(i|\boldsymbol{\omega}) = 0,$$

which guarantees the satisfactory of the complementary slackness condition for inequality constraint (23). Thus, KKT conditions hold, which means that $\{p(i|\boldsymbol{\omega})\}$ are optimal.

References

Ben-Akiva, M.E., and Lerman, S., (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Boyacı, T., and Akçay, Y., (2018). Pricing when customers have limited attention. *Management Science*, 67(7):2995-3014.

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

Cover, T.M., and Thomas, J.A., (2006). *Elements of Information Theory (Second edition)*. Wiley-Interscience.

Dean, M., and Neligh, N, (2023). Experimental Tests of Rational Inattention. *Journal of Political Economy,* in press, doi: 10.1086/725174.

Du, L.L., Han, L.S., Li, X.Y., (2014). Distributed coordinated in-vehicle online routing using mixed-strategy congestion game. *Transportation Research Part B: Methodological*, 67(2014) 1-17.

Duncan, L.C., Watling, D.P., Connors, R.D., Rasmussen, T.K., and Nielsen, O.A., (2020). Path Size Logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with GPS data. *Transportation Research Part B: Methodological*, 135:1-40.

Fosgerau, M., Frejinger, E., and Karlstrom, A., (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B: Methodological*, 56:70-80.

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

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

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.

Huettner, F., Boyacı, T., and Akçay, Y., (2019). Consumer choice under limited attention when alternatives have different information costs. *Operations Research*, 67(3):671-699.

Jaynes, E.T., (1957). Information Theory and Statistical Mechanics. *Physical Review*, 106(4):620-630.

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

Knies, A., Lorca, J., and Melo, E., (2022). A recursive logit model with choice aversion and its application to transportation networks. *Transportation Research Part B: Methodological*, 155: 47-71.

Ma, J.S., and Fukuda, D., (2015). A hyperpath-based network generalized extreme-value model for route choice under uncertainties. *Transportation Research Part C: Emerging Technologies*, 59:19-31.

Mai, T., Fosgerau, M., and Frejinger, E., (2015). A nested recursive logit model for route choice analysis. *Transportation Research Part B: Methodological*, 75:100-112.

Mai, T., (2016). A method of integrating correlation structures for a generalized recursive route choice model. *Transportation Research Part B: Methodological*, 93:146-161.

Mai, T., Yu, X., and Gao, S., (2021). Routing policy choice prediction in a stochastic network: Recursive model and solution algorithm. *Transportation Research Part B: Methodological*, 151:42-58.

Matějka, F., (2016). Rational inattentive seller: Sales and discrete pricing. *Review of Economic Studies*, 83(3):1125-1155.

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

Papola, A., Tinessa, F., and Marzano, V., (2018). Application of the combination of random utility models (CoRUM) to route choice. *Transportation Research Part B: Methodological*, 111:304-326.

Ravid, D., (2020). Ultimatum bargaining with rational inattention. *American Economic Review*, 110(9):2948-2963.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27 (3):379-423.

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

Sims ,C.A., (2006). Rational inattention: Beyond the linear-quadratic case. *American Economic Review*, 96(2):158-163.

Sims, C.A., (2010). Rational inattention and monetary economics. *Handbook of Monetary Economics*, 3:155-181.

Simon, H.A., (1955). A behavioural model of rational choice. *The Quarterly Journal of Economics*, 69(1):99-118.

Steiner, J., Stewart, C., and Matějka, F., (2017). Rational inattention dynamics: Inertia and delay in decision making. *Econometrica*, 85(2):521-553.

Walker-Jones, D., (2023). Rational inattention with multiple attributes. *Journal of Economic Theory*, 212:105688.

Wang, R.X., (2021). Discrete choice models and applications in operations management. *INFORMS TutORials in Operations Research*, null(null):199-226.