International Journal of Innovative Computing, Information and Control
Volume **16**, Number **4**, August **2020**

ICIC International ©2020 ISSN 1349-4198
pp. 1323–1336

TRAVEL MODE CHOICE MODELING VIA INFERENCE DIAGRAM CONSIDERING TRAVELERS' TASTE OF RISK

Xingpei Yan[1,2] and Lang Wei[1]

[1]School of Automobile
Chang'an University
Middle-Section of Nan'er Huan Road, Xi'an 710064, P. R. China
qch_1@chd.edu.cn

[2]Department of Transport Policy and Planning Research
Road Traffic Safety Research Center of the Ministry of Public Security
B#14, New World Center Office Building, Chongwenmenwai Avenue
Dongcheng District, Beijing 100062, P. R. China
yanxp1989@126.com

Received November 2019; revised March 2020

Abstract. *Current research approaches have not adequately considered the potential uncertainties of different modes. This paper develops an inference diagram (ID) approach to model travel mode choice with the consideration of travel time/cost uncertainty. The proposed ID model captures probabilistic relations between a trip's uncertain attributes and certain attributes. Given a trip's certain attributes, the uncertain attributes (e.g., travel time) can be inferred based on the Bayesian theorem; and the expected utility (EU) theorem is utilized for choice decision making. Simulation-based optimization (SBO) is applied to estimating the parameters of the ID model. There are two major contributions in this paper: 1) the ID model considers uncertain attributes inference in the decision-making process of travel mode choice; 2) the ID model is capable of considering travelers' heterogeneous tastes of risk. A real-world travel survey data conducted in the Washington D.C. area is used for a case study. We construct ID models with different tastes of risk; a traditional Logit model is fitted for comparison. The results indicate the ID model is superior in classification accuracy compared with MNL models. With travelers' heterogeneous tastes of risk, the ID model is capable of providing interesting findings in travel behavior research.*

Keywords: Travel mode choice, Bayesian statistics, Travel uncertainty, Taste of risk

1. **Introduction.** Travel mode choice is an essential topic in travel demand and urban mobility analysis. Accurate classification of travelers' mode choice is meaningful for various applications, such as travel demand forecast, congestion mitigation, and transportation system management. There have been a variety of approaches proposed in this research field [1-4]. These models, which can be classified into parametric models (i.e., random utility (RM) models) and non-parametric models (i.e., machine learning models), discussed in detail about the mathematic formulation and parameter estimation.

For decades, the multinomial Logit (MNL) model and its family have been widely applied to travel mode choice problems. MNL was first developed by McFadden, which assumes travelers tend to take the mode with the maximal RU [6]. RU is usually formulated as a linear sum of general costs associated with trip attributes (e.g., travel time and fare) and social-demographical variables (e.g., income level, gender). On one hand, the estimation of model parameters can be helpful to understand the relative importance

DOI: 10.24507/ijicic.16.04.1323

of different travel and social-demographical attributes in decision-making. On the other side, it considers the hidden causal factors that are not included in the utility function as one mode-specific constant. MNL also assumes the modes are not correlated with each other, which is referred to as the independence of irrelevant alternatives (IIA). By adding an unobserved component of utility correlated among modes, nested Logit (NL) model was developed to overcome the IIA issue [7,8]. The model structure of MNL family becomes more and more complex to represent the statistical relation between modes and attributes of modes or even attributes of a traveler. The big MNL family (also referred as RU models) includes cross-nested Logit model [9], generalized extreme value model [1], Bayesian NL model [10], continuous cross-nested Logit models [11], mixed Logit model [12], latent class model [3,13], etc.

Another approach of travel mode choice modeling is machine-learning. Unlike the MNL family with a decision-making theorem (i.e., RU maximization), machine-learning approaches attempt to build variables connections via some algorithms. The focus on data itself and the flexibility in model structure make it possible to offer different insights compared with the MNL family [14]. There are several machine-learning models frequently seen in the field of choice modeling, e.g., decision tree model [2,14,15], neural network model [5,16], support vector machine [17], Bayesian network model [4]. These models can handle large-size databases, saving much time in estimation and finding more complex relations compared with discrete choice models. Yet in the field of traffic behavior, machine-learning methods can hardly capture the key factors that travel behavior scientists usually concern, e.g., value of time (VOT). Another recognized issue is that machine-learning models are sensitive to training data, leading to biased estimations in the sense of insufficient or biased samples.

It is undeniable that the aforementioned models have made great progress in travel choice modeling. These approaches commonly assumed that travelers make decisions with a certain knowledge on trip attributes. In other words, the probability of choosing a particular mode is "some function of the attributes of the modes and a set of socio-demographic variables" [18]. The validity of this assumption can be denied because, in reality, the exact attributes of a mode can only be obtained after experiencing. Namely, a traveler can hardly know the exact travel time and fare of driving mode before he/she has already decided to drive and finished the trip. There have been a number of uncertainty-based travel choice studies in the literature. Jou and Kitamura used decision rules to frame possible gain and loss in terms of the earliest acceptable arrival time and official work start time [19]. Polak et al. considered travel time variability on mode/departure time choice by using the expected value [20]. De Palma et al. summarized both expected utility (EU) models and RU models under uncertainty and claimed the importance of linking risk and uncertainty with travel choice modeling [21]. Currently, most of these studies have considered departure time choice and route choice [18,32,33]; while such considerations were not frequently seen in mode choice. The lack of uncertainty-based mode choice study is due to the lack of multi-modal mobility data. Namely, traditional household surveys record detailed multi-modal trip dairies but do not contain reliability-related information.

This paper aims to enhance the modeling of uncertainty-based travel choice by developing an inference diagram (ID) based mode choice model. An ID belongs to the interdisciplinary family of causal models and graphical models. It has been applied in different research fields, such as environmental policy decision making [22], and business decision [23]. The proposed ID constructs dependency relations between uncertain attributes (e.g., travel time and cost) and certain attributes (e.g., trip distance and social-demographical variables) via the Bayesian theorem [24]. In the ID model, travelers evaluate different modes by inferring the corresponding uncertain attributes based on conditional probability and make choices under the EU theorem [25]. Both EU and the aforementioned RU theorems reflect decision-making mechanism in psychology. Unlike RU models which utilize an error term to model the uncertainty of choosing a mode, the ID model treats uncertain attributes as random variables. The proposed ID model also keeps the capability to derive key travel demand factors (e.g., VOT), and it can model travelers' heterogeneous tastes of risk. Taking the empirical probability distributions of multi-modal travel times, the ID model can work with traditional household travel surveys for behavior analysis. To test the ID decision-making mechanism, we fit the model with the 2007/2008 Transportation Planning Board – Baltimore Metropolitan Council Household Travel Survey (TPB-BMC HHTS) data. The coefficients of the ID model are estimated via simulation-based optimization (SBO). Comparisons with the MNL family are presented to provide insightful findings of ID models in travel choice classification.

The remainder of the paper is organized as follows. Section 2 presents a detailed introduction of the ID model, in terms of travelers' decision-making mechanism, model formulation, and coefficients estimation. Section 3 provides the 2007/2008 TPB-BMC HHTS data and the underlying ID mode choice models with different tastes of risk. Section 4 shows the results of the ID models, as well as the comparisons against an MNL model (as a benchmark). Finally, Section 5 concludes the paper and discusses future research issues.

2. Decision Making under Inference Diagram.

Bayesian decision-making theory allows researchers to characterize the probabilistic relation between a decision and its related prospection (with both known and unknown information). A travel decision can be modeled via Bayesian models because the experience of a trip can be formulated as a combination of observed and unobserved attributes. In transportation engineering, a number of research efforts have been conducted towards the development of Bayesian travel decision mechanism [4,14,32]. These studies analyzed the influences of unobserved variables on travelers' decisions, but the magnitude of travel time reliability has not been considered. Empirical evidence has shown the importance of reliability on travelers' daily choices [32]. A traveler's decision-making mechanism under uncertainty can be built based on the ID model, which follows a general way of how people make decisions.

2.1. Introduction of inference diagram.

An ID can be defined as $N = (\{X\}, \{E\}, \{P\}, U)$. $\{X\}$ represents the set of nodes (i.e., variables). There are four types of nodes in an ID: decision nodes, uncertain nodes, deterministic nodes, and utility nodes. Decision nodes refer to the decision to make, which are categorical variables. Deterministic nodes represent the quantified attributes that decision-makers have already known before making choices; on the contrary, uncertain nodes represent the attributes that are not known until a decision has been made. Utility nodes represent the value of utility. $\{E\}$ represents the set of edges. There are three types of edges corresponding to the to-node of the edge: 1) if the to-node is a decision node, the edge is called information edge, indicating that the information can be obtained before making decisions; 2) the edge is a conditional edge once the to-node is an uncertain or a deterministic node, and this means the attribute of the from-node depends on the to-node; 3) if the to-node is a utility node, the edge is a functional edge, indicating the from-node is a component of the utility function. $\{P\}$ denotes the set of conditional probability distributions/tables corresponding to each conditional edge. U denotes the utility function.

Figure 1 is an illustrative example of how an ID models the decision-making process of whether to drive or to take the subway. There are four nodes: X_T denotes the roadway

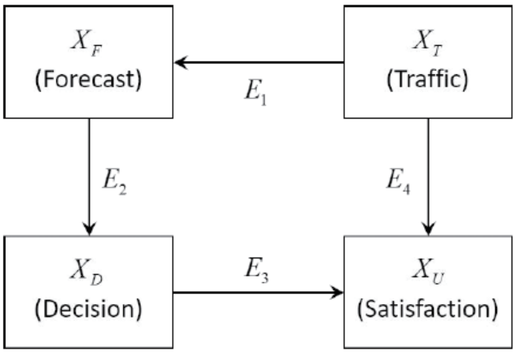

FIGURE 1. An illustration: Subway taking decision under ID

traffic conditions (congested or not), which is an uncertain node; X_F denotes the forecast of traffic congestion (congested or not), which is a deterministic node because we assume the person knows the traffic forecast before he/she leaves home; X_D is a decision node that denotes a person's decision (driving or not); and X_U is a utility node, representing the satisfaction (utility) of this person after leaving home. The four edges E_1, E_2, E_3, and E_4 basically explain the decision-making process. The transportation monitoring center will predict near-future traffic conditions based on historical and real-time traffic data (edge E_1). The person will decide whether or not to take the subway based on the forecast (edge E_2). Once the decision has been made, he/she will leave home and his/her satisfaction (utility) is determined both by the congestion level and by his/her decision (edges E_3 and E_4).

2.2. **Expected utility theorem and decision-making in an inference diagram.** In an ID model, the quantified satisfaction is represented by the utility, which has been used in aforementioned choice models. For simplicity, we assume a linear utility structure for node X_U in our subway choice example in Figure 1.

$$U = \alpha_1 \cdot X_D + \alpha_2 \cdot X_T \tag{1}$$

where α_1 and α_2 are coefficients of the decision and congestion state, respectively.

This is a normal utility function under which the decision-maker is assumed to be risk-neutral. Nonlinear transformations of the uncertain attributes can be used to account for different tastes of risk:

$$U = \alpha_1 \cdot X_D + \alpha_2 \cdot g(X_R, \varphi) \tag{2}$$

where $g(\cdot)$ is a nonlinear function. In this study, we set $g(\cdot)$ as a power function X^{φ}, where X is the attribute of the uncertain node (e.g., roadway traffic condition) and φ is the power. If φ is greater than 1, $g(\cdot)$ is concave and the expected attribute of an uncertain prospect will be greater than the objective expected attribute, implying a risk-averse behavior; on the contrary, a φ smaller than 1 indicates a risk-seeking behavior; while φ equal to 1 means a risk-neutral behavior.

One strong assumption in this behavior modeling process is that the person's satisfaction can be quantified once he/she knows the congestion level and the choice. In other words, the utility can be determined right after X_D and X_T are observed. Under this framework, we use EU theorem to model the decision-making process. In decision theory, the EU theorem describes how people make choices under risky (probabilistic) outcomes of different alternatives [25]. Decision-makers will choose the alternative which can maximize the expected value of the utility function with respect to all possible outcomes. The EU theorem assumes decision-makers have a subjective probability measure of the uncertain traffic condition node X_T. The EU under decision $X_D = i$ can be expressed as

$$E\left(U|X_D = i\right) = \sum_j U\left(X_T = j_T, X_D = i\right) \cdot \Pr\left(X_T = j_T|X_D = i, X_F = j_F\right) \tag{3}$$

where $\Pr(X_T = j_T|X_D = i, X_F = j_F)$ is the probability of the roadway traffic condition j_T given the decision i and congestion forecast j_F. In our example, X_D, j_F and j_T can only be "yes" or "no" (i.e., binary variable), and since the roadway traffic condition is independent of the decision, we have

$$\Pr(X_T = j_T|X_D = i, X_F = j_F) = \Pr(X_T = j_T|X_F = j_F) \tag{4}$$

In a more general case, the EU under decision $X_D = i$ is

$$E\left(U|X_D = i\right) = \sum_J \Pr\left(X^R = J|X_D = i\right) U\left(X^R = J, X_D = i\right) \tag{5a}$$

$$X^R = \left\{X_1^R, \ldots, X_k^R\right\} \tag{5b}$$

$$J = \{j_1, \ldots, j_k\} \tag{5c}$$

where X^R is a set of nodes that contribute to the utility function, and J is the vector of the outcomes of these variables. The key difference between EU maximization and RU maximization is that the former one utilizes probability to represent the uncertainty of utility, while the latter one has a random error term in the utility function.

Once the utility function is determined, the calculation of EU requests the conditional probability of the set $\left\{X_1^R, \ldots, X_k^R\right\}$ under X_D. Given the value of the deterministic nodes and conditional edges, the conditional probability table can be estimated via probabilistic inference [26], which releases the IIA assumption. More details about probabilistic inference for the mode choice model are presented in Section 3.2. The structure of the conditional edges, as well as the ID graph, can be predetermined by behavior and information dependency assumptions with respect to the research issue. Thus, the left task is to estimate the coefficients of the utility functions and tune the optimal coefficients to maximize the classification accuracy. In this study, we apply SBO to estimating the optimal coefficients. The construction of an ID model can be formed using the following steps.

1) Determine decision nodes, deterministic nodes, and uncertain nodes in a database.
2) Define the dependent relationship between the uncertain nodes and other nodes, and draw the conditional edges between these nodes.
3) Derive the conditional probability distribution of the uncertain nodes from the training database via probabilistic inference. This conditional probability distribution is regarded as the decision-makers' subjective distribution of the uncertain variables.
4) Determine the nodes in the utility function, and then build the functional edges. Formulate the utility functions under different decision scenarios with related nodes. In this study, we formulate the mode choice utility functions based on previous mode choice RU models [7], which assumes the utility is a linear combination of travel time, travel cost, and other variables.
5) Use SBO to estimate the optimal coefficients of the utility function that maximize the classification accuracy with respect to the training dataset.

2.3. **Estimating inference diagram utility coefficients via simulation-based optimization.** SBO is a statistic based optimization method, which deals with optimization problems without an analytical solution [27-30]. "Simulation-based" means the value of the objective function (Y) can only be obtained from a simulation (or a black box) output. In the ID model, the relation between the classification accuracy and the coefficients of the utility function is not derivable directly through mathematical formulations. Therefore, we use SBO such that the coefficients of the utility function are the decision variables (A) and the classification accuracy is the objective value (Y). Given A, the choice for each traveler can be estimated based on the ID model (i.e., our "simulation platform"). The objective is to find the optimal coefficients of the utility function to maximize the classification accuracy among the training dataset (the number of correct classifications divided by the training sample size).

As illustrated by the flowchart in Figure 2, we utilized the surrogate-based optimization method [27-30] to solve this ID utility coefficient estimation problem. Firstly, N initial coefficient vectors $(A_1, A_2, \ldots, A_N)$ are generated via the Latin hypercube sampling method. Then, the classification accuracy of each initial coefficient vector will be evaluated in the ID model. That is, for each traveler in the database, we calculate the EU under each travel mode alternative and assume the individual chooses the travel mode that can maximize his/her EU. The classification accuracy is obtained by comparing the mode choice outputs from the ID model and the actual choices in the survey database. After estimating N initial vectors, the relation between A and Y is fitted via the radial basis function method and ordinary Kriging method. The leave-one-out cross-validation is conducted to measure the goodness-of-fit. Based on the fitted response surface of A and Y, a new infill coefficient vector with a high probability to lead to a higher objective value is generated and selected. The new infill coefficient vector is evaluated via ID and all the initial and infill vectors are used to fit a new surrogate model. The infill process would not stop until the total number of infill vectors reaches M. Finally, we select the vector with the highest classification accuracy among the $N + M$ samples to be the optimal solution. Readers can refer to [27-30] for a detailed introduction of surrogate-based optimization.

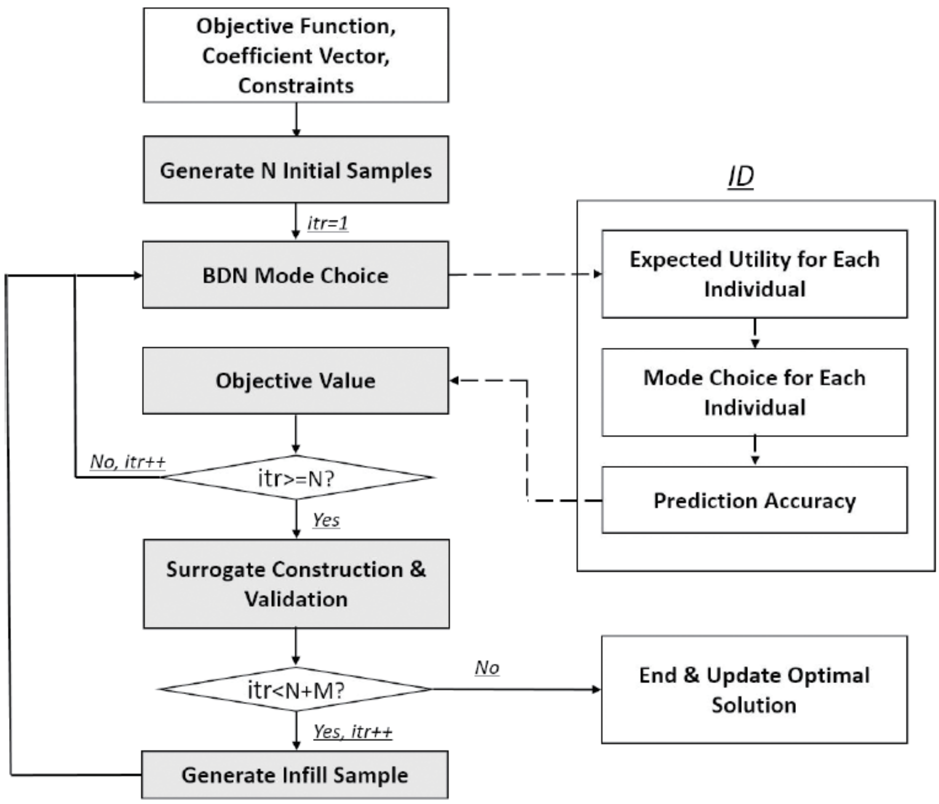

FIGURE 2. Flowchart of SBO for coefficients estimation

3. **Modeling Travel Mode Choice via Inference Diagram.** This section presents the modeling of travel mode choice via ID. The survey dataset and the variable selection are introduced. This section also discusses the ID model's structure, formulation of utility functions, and the setup of SBO.

3.1. **2007/2008 travel survey dataset.** The 2007/2008 TPB-BMC HHTS dataset is used for mode choice modeling in this study, which includes about 14,000 households in the Washington and Baltimore regions. Randomly selected households were asked to complete a travel diary that documented the activities of all household members on an assigned day. Various information was recorded including trip, person, household, and vehicle information. The Metropolitan Washington Council of Governments (MWCOG) travel forecasting model is used to generate travel time and cost information for different travel modes. Travel time of transit, driving, and carpool is provided by travel time skim matrices from the MWCOG model. Travel time of walk/bike is estimated using reported trip distance divided by speed. The speed is exogenously defined as 4.87 miles per hour. For travel cost, the zone-to-zone transit fare files are used for transit. In the MWCOG model, travel cost for driving includes two parts: parking cost and auto operating cost. Parking cost is provided by zone level parking cost file while auto operating cost is assumed as 10 cents per mile. Travel cost of carpool is assumed to be half of driving cost.

In this study, four travel modes are considered: transit, driving, carpool and walk/bike. Explanatory variables used in mode choice modeling include income, number of vehicles,

Table 1. Nodes (variables) considered in the ID mode choice model

Node	Value
Decision Node	
Travel mode (Mode)	Discrete variable:
	0 for car driver
	1 for car passenger
	2 for transit
	3 for bike/walk
Deterministic Nodes for Social-Demographical Attributes	
Income (Income)	Continuous variable
	Annual salary in thousand dollars
Household Vehicle Num. (NHV)	Discrete variable
Household Bike Num. (NHB)	Discrete variable
Deterministic Nodes for Certain Travel Attributes	
Trip Distance	Continuous variable
Dummy Distance 1 (Dis1)	Discrete variable:
	0 for trip distance < 8.0 miles
	1 for trip distance ≥ 8.0 miles
Dummy Distance 2 (Dis2)	Discrete variable:
	0 for trip distance ≥ 0.6 miles
	1 for trip distance < 0.6 miles
Uncertain Nodes for Uncertain Travel Attributes	
Car Travel Time ($\mathrm{TT_{Car}}$)	Continuous variable (minutes)
Transit Travel Time ($\mathrm{TT_{Tran}}$)	Continuous variable (minutes)
Walk and Bike Travel Time ($\mathrm{TT_{WB}}$)	Continuous variable (minutes)
Car Travel Cost ($\mathrm{TC_{Car}}$)	Continuous variable (US dollar)
Transit Travel Cost ($\mathrm{TC_{Tran}}$)	Continuous variable (US dollar)

number of bikes, trip distance, travel time and cost, as summarized in Table 1. Generally, the level-of-services variables for different modes (e.g., travel times and travel costs) are the essential variables in travel choice analyses and social-demographical variables help one to discover the heterogeneities in travelers' preferences [6-14]. Two binary distance variables "Dis1" and "Dis2" are created: if the trip distance is over 8.0 miles, "Dis1" is true; if the trip distance is below 0.6 miles, "Dis2" is true. We use these two binary distance variables to examine whether there is a preference to transit when trip distance is over 8.0 miles, or preference to walk/bike when trip distance is within 0.6 miles. Trips that have access to transit within TPB modeled area are used for modeling (61,615 records). In order to study the generalization performance of the ID model, the sample is randomly divided into two parts: 60 percent of the data for training (36,969 records) and 40 percent of the data for testing (24,646 records).

3.2. **Inference diagram mode choice model.** The structure of this ID model is shown in Figure 3. The values of "Dis", "Dis1", "Dis2", "Income", "NHV", and "NHB" are available before the decision so that the corresponding nodes are deterministic nodes. Each of these deterministic nodes has an information edge pointing to the decision node "Mode", which is not shown in Figure 3. The values of nodes "TT" and "TC" are uncertain, while a traveler has subjective beliefs with respect to their conditional probability distributions. The subjective distribution of "TT" is conditional on "Dis" and "Mode"; similarly, the distribution of "TC" is conditional on "Dis", "Mode", and "TT".

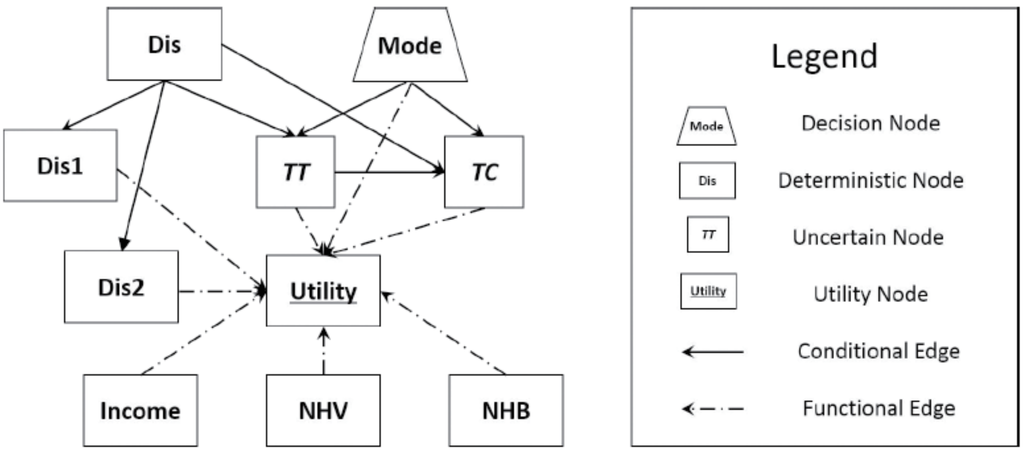

FIGURE 3. ID structure for mode choice

The conditional probability distributions of the uncertain nodes are estimated from the training database of the TPB/BMC survey via probabilistic inference, and we assume they follow linear conditional Gaussian distribution.

$$f(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{6a}$$

$$\Pr(TT_i=j_{TT}|Dis=j_{Dis})=f\left(j_{TT};\theta_{i,0}^{TT}+\theta_{i,Dis}^{TT}\cdot j_{Dis},\delta_i^{TT}\right) \tag{6b}$$

$$\Pr(TC_i=j_{TC}|Dis=j_{Dis},TT=j_{TT})=f\left(j_{TC};\theta_{i,0}^{TC}+\theta_{i,Dis}^{TC}\cdot j_{Dis}+\theta_{i,TT}^{TC}\cdot j_{TT},\delta_i^{TC}\right) \tag{6c}$$

where $f(x;\mu,\sigma)$ is the probability density function of Gaussian distribution. Equation 6(b)/6(c) indicates that the expected value of node travel time/cost is a linear sum of the value the from-nodes of its conditional edges in Figure 3. $\theta_{i,0}^{TT}/\theta_{i,Dis}^{TT}$ denotes the parameters of mean travel time for mode i, and δ_i^{TT} denotes the standard deviation. Similarly, $\theta_{i,0}^{TC}$, $\theta_{i,Dis}^{TC}$, and $\theta_{i,TT}^{TC}$ denote the parameters of mean travel cost of mode i, and δ_i^{TC} denotes the corresponding standard deviation. Parameters θ and δ are estimated based on the training dataset via maximum-likelihood estimation (MLE), in which the log-likelihood functions for travel time and travel cost are

$$LL_i^{TT}(\theta,\delta) = \sum_l \log\left(\Pr_l(TT_i = j_{TT,l} | Dis = j_{Dis,l})\right) \tag{7a}$$

$$LL_i^{TC}(\theta,\delta) = \sum_l \log\left(\Pr_l(TC_i = j_{TC,l} | Dis = j_{Dis,l}, TT = j_{TT,l})\right) \tag{7b}$$

where l denotes the index for travelers.

The utility functions, which consider all the nodes that have a functional edge to node "Utility", are illustrated below.

$$U_0 = \alpha_{Drive} + \beta_{Car} \cdot g(TT_{Car}) + \gamma_{Drive} \cdot g(TC_{Car}) + \eta_{Car} \cdot NHV + \delta \cdot Income \tag{8a}$$

$$U_1 = \alpha_{Pass} + \beta_{Car} \cdot g(TT_{Car}) + 0.5 \cdot \gamma_{Pass} \cdot g(TC_{Car}) \tag{8b}$$

$$U_2 = \alpha_{Tran} + \beta_{Tran} \cdot g(TT_{Tran}) + \gamma_{Tran} \cdot g(TC_{Tran}) + \chi_{Dis1} \cdot Dis1 \tag{8c}$$

$$U_3 = \alpha_{WB} + \beta_{WB} \cdot g(TT_{WB}) + \eta_{Bike} \cdot NHB + \chi_{Dis2} \cdot Dis2 \tag{8d}$$

$$g(s) = s^{\varphi} \tag{8e}$$

where U_0, U_1, U_2, and U_3 are the utilities under travel modes "drive", "passenger", "transit" and "walk and bike", respectively; α_{Drive}, α_{Pass}, α_{Tran}, and α_{WB} denote the interceptions of the four modes; β_{Car}, β_{WB}, and β_{Tran} are the coefficients for travel time of different modes; here we assume the travel time coefficients for "Driver" and "Passenger" are the same, and the passenger cost is half of the driving cost; γ_{Drive}, γ_{Pass}, and γ_{Tran} are the coefficients for travel cost; η_{Car} and η_{Bike} denote the coefficients of household number of vehicles/bikes, respectively; χ_{Dis1} and χ_{Dis2} denote the coefficients of the binary distance for "transit" and "walk and bike"; δ denotes the coefficient of income; $g(\cdot)$ is the nonlinear transformation function defined in Section 2.2. In this paper, we have attempted different values for risk preference modelling (i.e., φ ranges from 0.5 to 2.0). Based on the classification accuracy with respect to different parameters, we consider three scenarios with φ equaling 0.8/1.0/1.2 to represent risk-seeking/risk-neutral/risk-averse travelers. Given the formulation, the EU under different travel modes can be calculated via Equations 5(a)-5(c), which can be summarized as

$$\begin{aligned} &E(U|Mode=i) \\ &= \iint \Pr\left(TT_i = j_{TT} | Dis = j_{Dis}\right) \Pr\left(TC_i = j_{TC} | Dis = j_{Dis}, TT = j_{TT}\right) U_i \mathrm{d}j_{TC} \mathrm{d}j_{TT} \end{aligned} \tag{9}$$

There are 15 coefficients to estimate in this ID mode choice model

$$A = (\alpha_{Drive}, \alpha_{Pass}, \alpha_{WB}, \alpha_{Tran}, \beta_{Car}, \beta_{WB}, \beta_{Tran}, \gamma_{Drive}, \gamma_{Pass}, \gamma_{Tran}, \chi_{Dis1}, \chi_{Dis2}, \delta)$$

We assume the travel time coefficients (β) and the travel cost coefficients (γ) to range from -10 to 0. The rest of the coefficients are assumed to have a positive impact on the utility, and range from 0 to 10. Since the range of coefficients has an equal length of 10, we used 0-1 mapping for each continuous variable to normalize all the variables in the training data set

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{10}$$

where x is the original value in the data set, $x_{\min}$ and $x_{\max}$ are the minimal and maximal values in the training set, and x' is the normalized value. For the validation data set, we still applied the same mapping for consistence. Under the domain of the coefficient vector space, we generated 10,000/2,000 of initial/infill coefficient vectors in the SBO process.

4. **Result Analysis.** In order to make a fair comparison with traditional RU models, we use the same nodes (variables) and the same formulation of utility functions to estimate an MNL model. The MNL model is estimated via MLE. The MNL (also ID) utility specification in Equation (8) is the final formulation after several adjustments until we get significant estimations in the MNL. This is because the ID framework does not have significant tests, so we rely on the estimation of an MLE to determine the significant explanatory nodes in the ID models. During the SBO process, the initial 10,000 vectors result in the best classification accuracy of around 67% in the three ID models; and the infill process improves the classification accuracy to around 73%.

The estimated coefficients and classification accuracy are shown in Table 2, in which each column represents the coefficients of a specific model. In terms of classification accuracy, the ID models tend to perform better than the MNL model in both the training and the test datasets. This is mostly because the SBO is seeking the coefficients with the highest classification accuracy of the training data. With different behavior modeling mechanisms and psychological assumptions, there is much difference in the coefficients of the utility functions and we can explain mode choice behavior in over one way. One reason for the differences in coefficients is because the inferred travel time and cost used in the ID models are different from the surveyed travel time and cost used in the MNL model. Another reason lies in the differences in the decision-making mechanism under utility. The MNL model presents decision as a probability because the utility function includes an error term such that:

$$\Pr(Mode = i) = \frac{e^{A_i \cdot X}}{\sum\limits_{ii=0}^{3} e^{A_{ii} \cdot X}} \tag{11}$$

where A_i is the coefficients under mode i, X is the vector of travel and social-demographical attributes. While the ID models use a dominating rule, since the uncertainty has already been considered in the EU function:

$$Mode = \underset{i \in \{0,1,2,3\}}{\arg\max}(E(U|Mode = i)) \tag{12}$$

In order to verify the superiority of proposed ID model on classifying mode choice, receiver operating characteristic (ROC) curve is shown in Figure 4. Travelers that are classified to drive (based on Equation (12)) are called positive predictions. Among the positive predictions, the true predictions are the correct ones, while the false predictions are the wrong ones. The ROC curve is plotted by the true positive rate against the false positive rate, such that the closer it is to the left upper corner, the better the classification is. Figure 4 indicates that all the ID models are more superior in identifying a driver than the MNL model. Moreover, the population in the survey area can be recognized as risk-neutral since the dashed black (risk-neutral) curve gives the best performance.

The intercepts in an MNL model reflect the preferences of different modes; while the intercepts in an ID model are only weighting factors to improve the classification accuracy. Thus, the highest intercept of driving in the ID models does not necessarily mean travelers have a preference for driving. On the contrary, the intercept has to be higher to guarantee the classification accuracy, since the majority of travelers in the survey chose driving. We conclude the intercepts in MNL and ID are incomparable. We believe the results in MNL indicate a preference in driving, which has the highest intercept. In addition, an MNL model tends to have higher coefficients in travel time and cost than the ID models; while the coefficients of social-demographical attributes are lower than the ID models. This can be because the information loses in travel time in ID, such that the social-demographical attributes become more important in decision-making.

TABLE 2. Model results

	MNL	ID 1.0 risk-neutral	ID 1.2 risk-averse	ID 0.8 risk-seeking
Intercept (Drive)	2.76 (**)	1.74 (**)	2.54 (**)	2.51 (**)
Intercept (Passenger)	2.53 (***)	1.41 (**)	0.25 (**)	0.72 (**)
Intercept (W & B)	0.17 (*)	0.00 (**)	1.05 (**)	0.75 (**)
Intercept (Transit)	0.00	1.43 (**)	1.84 (**)	1.54 (**)
TT_{Car}	−2.10 (**)	−1.51 (**)	−1.11 (**)	−2.26 (**)
TT_{WB}	−12.4 (**)	−5.41 (***)	−6.93 (***)	−9.78 (**)
TT_{Tran}	−2.01 (*)	−1.27 (*)	−1.39 (*)	−1.02 (*)
TC_{Drive}	−6.87 (***)	−4.89 (**)	−6.78 (**)	−5.47 (***)
$TC_{Passenger}$	−17.62 (**)	−9.15 (***)	−7.94 (**)	−8.82 (***)
TC_{Tran}	−0.12 (*)	−0.46 (*)	−2.64 (**)	−1.45 (*)
NHV	3.70 (*)	9.84 (*)	6.11 (*)	7.12 (*)
NHB	1.24 (**)	2.21 (**)	1.64 (*)	2.17 (**)
Income	0.09	0.23 (*)	1.63 (*)	0.68 (*)
Dis1	0.41 (**)	0.34 (**)	0.14 (**)	0.03 (**)
Dis2	2.56 (***)	3.02 (**)	1.84 (***)	2.34 (**)
Classification Accuracy of Training Data	65.3%	75.2%	75.0%	74.8%
Classification Accuracy of Test Data	62.8%	70.6%	70.5%	70.3%

Significance Levels: ***0.01, **0.05, *0.1. The significance level for the ID model is based on the statistical tests with 200 estimations.

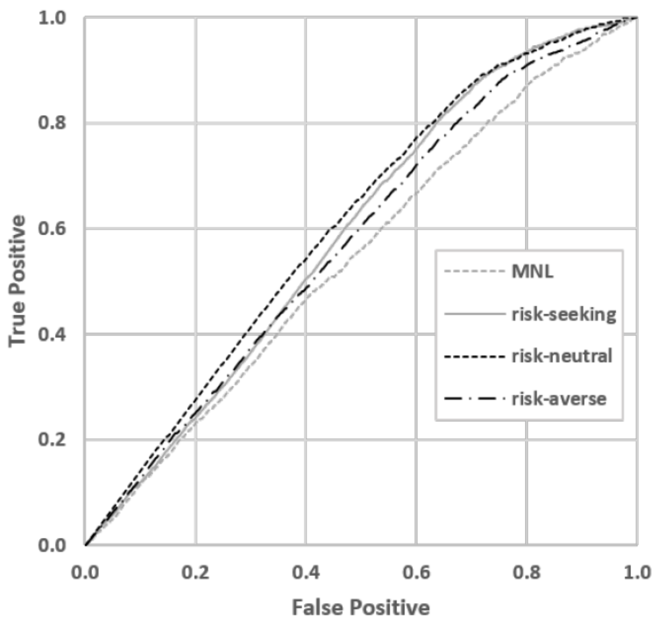

FIGURE 4. ROC curve of mode driving

Comparing MNL with ID 1.0, we find that the driving VOTs derived from the two models are 18.3 and 18.5 dollars per hour respectively (the VOT is calculated via $60 \cdot \beta_{Car}/\gamma_{Drive}$). This indicates the relative importance between travel time and travel cost may not have a big change when both travel attributes become uncertain in the mode choice process (ID 1.0). We also observe some interesting findings when comparing ID 1.0 (risk-neutral) with ID 0.8 (risk-seeking) and ID 1.2 (risk-averse): in the risk-seeking scenario, the coefficient of driving travel time is larger than that in the risk-neutral scenario, while the coefficient of transit travel time is smaller. This means once we assume travelers to be risk-seeking, they tend to have higher driving VOT (24.8 dollars per hour) than risk-neutral people; however, travelers still decide to drive because driving is risker than the other modes. On the contrary, if we assume the population to be risk-averse, they tend to be "driving lovers" (driving VOT is 9.8 dollars per hour), so that the majority still choose driving even though it is risky. Since it is unrealistic to assume all the travelers to be homogeneous in the taste of risk, we can either use technics such as latent class models to estimate different tastes of risk among travelers.

To summarize, this case study is the beginning of using ID to model travel choice. We simply use structure ID and linear utility functions at this stage to compare it with an RU based MNL model. The results in terms of classification accuracy indicate the ID model with an SBO estimation performs better than MNL model. More work is expected to better understand the influence of model structure under ID. For instance, we may include more social-demographical and traffic attributes, try nonlinear utility functions, use decision theorem that integrates RU and EU, or consider travel time uncertainty in MNL.

5. Conclusions and Discussions.

This paper developed an ID-based travel mode choice model to incorporate travel time and cost uncertainty into travel decision-making process. An ID model is integration between probability model, graphical model, and EU based decision-making theorem. The proposed ID model estimates probabilistic dependency relations, as travelers' subjective beliefs, between travel time and cost and trip distance via Bayesian probabilistic theorem. Based on the subjective conditional distributions, travelers inference the values of travel time and cost and then make choices via the EU theorem. SBO is used to estimate the coefficients in the utility functions, which are the linear sum of travel attributes and social-demographical attributes. Travelers' taste of risk is modeled via a nonlinear transfer of travel attributes.

A regional survey data was used to construct ID models with different tastes of risk (risk-neutral, risk-seeking and risk-averse), as well as an MNL model for comparison. The ID models provide higher classification accuracy than the MNL model. In risk-neutral scenario, the relative importance between travel time and cost may not have a significant difference whether or not the travel attributes are uncertain. In the risk-seeking scenario, we find the attitude for driving is not as favorable as in a risk-neutral assumption; while under the risk-averse assumption, we obtain opposite conclusions. Since it is unrealistic to assume a homogeneous taste of risk among the whole population, we will consider heterogeneous taste of risk in future research. Nonlinear utility functions will be tested for complex correlations between traffic attributes and the quantified satisfaction, e.g., the formulation of utility functions may vary across subgroups of travelers. Moreover, some other uncertainty-based decision theories such as prospect theory [8], regret theory [31], and the integration between RU and EU can also be incorporated with ID models. We will also incorporate the proposed methodology into other transportation problems, such as pricing and channel coordination [34] and demand management [35].

Acknowledgment. The research is supported by Central Public-Interest Scientific Institution Basal Research Fund (Grant No. 111041000000180001190103). We all acknowledge Dr. Zheng Zhu, who gave us helpful suggestions on this research task.

REFERENCES

[1] K. A. Small, A discrete choice model for ordered alternatives, *Econometrica: Journal of the Econometric Society*, pp.409-424, 1987.

[2] G. Wets, K. Vanhoof, T. Arentze and H. Timmermans, Identifying decision structures underlying activity patterns: An exploration of data mining algorithms, *Transportation Research Record: Journal of the Transportation Research Board*, vol.1718, pp.1-9, 2000.

[3] E. Ben-Elia, M. Bierlaire and D. Ettema, A behavioural departure time choice model with latent arrival time preference and rewards for peak-hour avoidance, *European Transport Conference*, Glasgow, Scotland, 2010.

[4] Z. Zhu, X. Chen, C. Xiong and L. Zhang, A mixed Bayesian network for two-dimensional decision modeling of departure time and mode choice, *Transportation*, vol.45, pp.1499-1522, 2018.

[5] D. A. Hensher and T. T. Ton, A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice, *Transportation Research Part E: Logistics and Transportation Review*, vol.36, no.3, pp.152-172, 2000.

[6] D. McFadden, Conditional logit analysis of qualitative choice behavior, in *Frontiers in Econometrics*, P. Zarembka (ed.), New York, NY, Academic Press, 1973.

[7] J. De Dios Ortuzar, Nested logit models for mixed-mode travel in urban corridors, *Transportation Research Part A: General*, vol.7, no.4, pp.283-299, 1983.

[8] F. S. Koppelman and C.-H. Wen, Alternative nested logit models: Structure, properties and estimation, *Transportation Research Part B: Methodological*, vol.32, no.5, pp.289-298, 1998.

[9] P. Vovsha, Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area, *Transportation Research Record: Journal of the Transportation Research Board*, vol.1607, pp.6-15, 1997.

[10] D. J. Poirier, A Bayesian analysis of nested logit models, *Journal of Econometrics*, vol.75, no.2, pp.163-181, 1996.

[11] J. Lemp, K. Kockelman and P. Damien, The continuous cross-nested logit model: Formulation and application for departure time choice, *Transportation Research Part B: Methodological*, vol.44, no.5, pp.646-661, 2010.

[12] M. Borjesson, Joint RP-SP data in a mixed logit analysis of trip timing decisions, *Transportation Research Part E: Logistics and Transportation Review*, vol.44, no.6, pp.1025-1038, 2008.

[13] C. Xiong, P. Hetrakul and L. Zhang, On ride-sharing: A departure time choice analysis with latent carpooling preference, *Journal of Transportation Engineering*, vol.140, no.8, 2014.

[14] C. Xie, J. Lu and E. Parkany, Work travel mode choice modeling with data mining: Decision trees and neural networks, *Transportation Research Record: Journal of the Transportation Research Board*, vol.1854, pp.50-61, 2003.

[15] L. Tang, C. Xiong and L. Zhang, Decision tree method for modeling travel mode switching in a dynamic behavioral process, *Transportation Planning and Technology*, vol.38, no.8, pp.833-850, 2015.

[16] H. Yang, R. Kitamura, P. P. Jovanis, K. M. Vaugh and M. A. Abdel-Aty, Exploration of route choice behavior with advanced traveler information using neural network concepts, *Transportation*, vol.20, no.2, pp.199-223, 1993.

[17] Y. Zhang and Y. Xie, Travel mode choice modeling with support vector machines, *Transportation Research Record: Journal of the Transportation Research Board*, vol.2076, pp.141-150, 2008.

[18] S. Rasouli and H. Timmermans, Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research, *Travel Behaviour and Society*, vol.1, no.3, pp.79-90, 2014.

[19] R. C. Jou and R. Kitamura, Commuter departure time choice: A reference-point approach, *Proc. of EWGT*, Bari, Italy, 2002.

[20] J. Polak, S. Hess and X. Liu, Characterising heterogeneity in attitudes to risk in expected utility models of mode and departure time choice, *The Transportation Research Board the 87th Annual Meeting*, Washington D.C., US, 2008.

[21] A. De Palma, M. Ben-Akiva, D. Brownstone, C. Holt, T. Magnac, D. McFadden and J. Walker, Risk, uncertainty and discrete choice models, *Marketing Letters*, vol.19, nos.3-4, pp.269-285, 2008.

[22] A. Sadoddin, R. A. Letcher, A. J. Jakeman and L. T. Newham, A Bayesian decision network approach for assessing the ecological impacts of salinity management, *Mathematics and Computers in Simulation*, vol.69, no.1, pp.162-176, 2005.

[23] D. P. Ames and B. T. Neilson, A Bayesian decision network engine for Internet-based stakeholder decision-making, *Proc. of ASCE EWRI Conf. – Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges*, 2001.

[24] T. Bayes and M. Price, An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, *Philosophical Transactions*, vol.43, pp.370-418, 1763.

[25] L. J. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 1947.

[26] G. E. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, 2011.

[27] X. Chen, L. Zhang, X. He, C. Xiong and Z. Li, Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network, *Computer-Aided Civil and Infrastructure Engineering*, vol.29, no.5, pp.359-381, 2014.

[28] X. Chen, Z. Zhu, X. He and L. Zhang, Surrogate-based optimization for solving mixed integer network design problem, *Transportation Research Record: Journal of the Transportation Research Board*, vol.2497, pp.124-136, 2015.

[29] X. He, X. Chen, C. Xiong, Z. Zhu and L. Zhang, Optimal time-varying pricing for toll roads under multiple objectives: A simulation-based optimization approach, *Transportation Science*, vol.51, no.2, pp.412-426, 2016.

[30] Z. Zhu, C. Xiong, X. M. Chen and L. Zhang, Calibrating supply parameters of large-scale DTA models with surrogate-based optimisation, *IET Intelligent Transport Systems*, vol.12, no.7, pp.642-650, 2018.

[31] P. C. Fishburn, Non-transitive measurable utility for decision under uncertainty, *Journal of Mathematical Economics*, vol.18, no.2, pp.187-207, 1989.

[32] Z. Zhu, S. Zhu, Z. Zheng and H. Yang, A generalized Bayesian traffic model, *Transportation Research Part C: Emerging Technologies*, vol.108, pp.182-206, 2019.

[33] Z. Zhu, X. Li, W. Liu and H. Yang, 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*, vol.131, pp.168-192, 2019.

[34] R. Zhang, J. Liu and B. Liu, Pricing decisions and channel coordination of fresh produce supply chain under multi-transportation modes, *ICIC Express Letters*, vol.12, no.1, pp.9-21, 2018.

[35] Q. Tang and X. Hu, Triggering behavior changes with information and incentives: An active traffic and demand management-oriented review, *The Evolving Impact of ICT on Activities and Travel Behaviour*, vol.3, pp.209-250, 2019.