Choice Models with Stochastic Variables and Random Coefficients

Mehek Biswas [a], mehekbiswas@iisc.ac.in

Chandra R. Bhat [b], bhat@mail.utexas.edu

Sulagna Ghosh [c], ghoshsulagna97@gmail.com

Abdul Rawoof Pinjari [a, d, *], abdul@iisc.ac.in

[a] Department of Civil Engineering, Indian Institute of Science (IISc), Bengaluru 560012, India

[b] Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, Austin, TX 78712, USA

[c] Indian Statistical Institute, Kolkata, 700108, India

[d] Centre for Infrastructure, Sustainable Transportation and Urban Planning (C*i*STUP), Indian Institute of Science, Bengaluru 560012, India

[*]Corresponding author

ABSTRACT

In travel choice models, variables describing alternative attributes such as travel time may have to be specified as stochastic because the analyst may not have accurate measurements of the attribute values considered by the decision-maker. Such stochasticity in alternative attributes is different from unobserved heterogeneity in the coefficients representing travellers' response to those attributes. Specifying only one of these as random while keeping the other fixed can potentially result in biased parameter estimates, inferior goodness-of-fit, and distorted information for policy analysis. Therefore, in this study, we propose an integrated choice and stochastic variable modelling framework with random coefficients (i.e., an *ICSV-RC* framework) that allows the analyst to accommodate stochasticity in alternative attributes and random coefficients on such attributes. In addition, we show that ignoring either source of stochasticity – stochasticity in alternative attributes or unobserved heterogeneity in response to the attributes – results in models with inferior goodness-of-fit and a systematic bias in all parameter estimates. We demonstrate this using simulation experiments for two different travel choice settings, one involving labelled mode choice alternatives and the other involving unlabelled route choice alternatives. In addition, we present an empirical analysis in the context of truck route choice to highlight the importance of accommodating both sources of variability – stochasticity in travel times and random heterogeneity in response to travel times.

Keywords: discrete choice, stochastic variables, random coefficients, identification, integrated choice and latent variable (*ICLV*) models

1 INTRODUCTION

In random utility maximization (RUM)-based travel choice models, it is common to assume that exogenous variables entering the utility functions are deterministic. Although the exogenous variables come from a distribution in the population, for any given observation in the data, it is common to build models considering a deterministic value (i.e., an observed value) that is a realization from the distribution of exogenous variables. In many situations, however, it may be more appropriate to specify the exogenous variables as stochastic, because the analyst may not be able to observe the specific realization of exogenous variables relevant to the observation. This is due to one or more of the following three reasons: (1) analyst's errors in measuring the true value of the variables, (2) travellers' perceptions of the values of the variables (that may be different from the measurements that the analyst may possess), and (3) inherent stochasticity in the variables, such as day-to-day variability. Each of these sources of stochasticity in explanatory variables is briefly discussed next in the specific context of travel time, which is a level-of-service variable used in travel choice models such as those for mode choice, route choice, and departure time choice.

The first reason for considering travel time entering the utility functions as a stochastic variable may be attributed to measurement errors by the analyst (Bhatta and Larsen, 2011*;* Ortúzar and Ivelic, 1987; Train, 1978). Such errors may arise due to: (a) the use of spatially aggregate (zone-to-zone) measures of level-of-service attributes instead of disaggregate, point-to-point measurements, (b) calculating travel times based on free flow speed assumptions instead of measuring actual speeds or other erroneous speed measurements, and (c) errors in coding of networks that result in erroneous travel time measurements. The second reason to treat travel time as a stochastic variable may originate from travellers perceiving travel times to be different from what may be objective travel times that the analyst might have measurements of (Daly and Ortúzar, 1990). The third reason is that travel time in transportation networks may be inherently stochastic due to day-to-day and intra-day variability in travel conditions on the network (Chen *et al*., 2011; Srinivasan *et al*., 2014; Biswas *et al*., 2019). As a result, travel time might follow a probability distribution making it difficult for the analyst to ascertain which of the different possible values of travel time was considered by the decision-maker in making the choice. In summary, all the three sources of stochasticity discussed above can potentially lead to error in the analyst's measurement of the precise duration of travel time considered by the traveller. One way to recognize such errors is to treat the travel time variable entering the utility functions of RUM-based choice models as a stochastic variable. [1]

As discussed in Diáz *et al*. (2015) and Ortúzar and Willumsen (2011), ignoring stochasticity in explanatory variables, if present, will, in general, lead to biased parameter estimates and distorted marginal rates of substitution (e.g., willingness to pay) during estimation. Important to note also is that some or all of the above-discussed stochasticity sources, while invoked in the specific context of travel time, can also apply to several other exogenous variables used in travel choice models. For example, crowding levels in transit modes can be stochastic in mode choice settings because of measurement errors. Travel costs measured by the analyst may be different from the costs travellers pay (or perceive) due to the different time scales in which the different costs occur (e.g., fuel costs are paid regularly, whereas insurance costs are paid once a year) and due to spatial aggregation of the travel locations.

Among the approaches used to accommodate stochasticity in variables in discrete choice models is the classic errors-in-variables (EIV) approach widely used in regression models (Stefanski and Carroll, 1985; Durbin, 1954; Gleser, 1981, etc.). Some choice modelling studies have adopted the EIV method through Rubin's multiple imputation (Rubin, 1987) for cases when data on exogenous variables are missing or unknown beyond certain interval bounds, such as for travel time (Steinmetz and Brownstone, 2005). Alternatively, studies such as Conniffe and O'Neil

[1] In this paper, we assume that the traveller (i.e., the decision-maker) attributes a single travel time value (duration) to a travel choice alternative, even in situations when travel times are inherently stochastic. The main issue dealt with in this paper is the analyst's difficulty in measuring the specific duration of travel time considered by the traveller, which can be any realization from the distribution of travel time known to the traveller. In this context, we ignore the possibility that the traveller might distort (or makes an estimation of) the actual distribution of travel time because of perception errors. There is a small but rich body of literature that accommodates how travellers deal with stochasticity in travel time (Liu and Polak, 2007; Polak *et al*., 2008; de Palma *et al*., 2007; de Palma *et al*., 2012). These studies draw from the literature on decision-making under risk and uncertainty, such as the classical expected utility theory (Von-Neumann and Morgenstern, 1947) and the prospect theory (Kahneman and Tversky, 1979; and Tversky and Kahneman, 1992). For example, Polak *et al*. (2008) use the expected utility theory where the traveller is assumed to work with the expected duration of travel time, but also incorporate variability in travel time through concepts of risk aversion. The reader is referred to papers by Rasouli and Timmermans (2014) and Li and Hensher (2011) for relevant reviews on these topics. Another approach is to empirically enhance the utility functions by including a measure of variability of travel time in addition to a central measure (e.g., average) of travel time (see, for example, Bhat and Sardesai, 2006; and Senbil and Kitamura, 2006). In the current study, however, we assume that the traveller associates a single travel time value with a given travel choice alternative. This assumption is not inconceivable, for emerging travel assistance/information sources provide likely durations of travel time for different travel choice alternatives available to the traveller. Admittedly, the current paper does not consider how travellers deal with stochasticity or variability in travel times, an area that needs attention in subsequent extensions of this work.

(2008) propose analytic expressions for estimators in the presence of missing data. Diáz *et al.* (2015) use the mixed logit approach to specify errors in variables as additional error components in the utility functions. A second method is the Integrated Choice and Latent Variable (*ICLV*) modelling approach (Ben Akiva *et al*., 2002; Alvarez-Daziano and Bolduc, 2013; Bhat and Dubey, 2014; Vij and Walker, 2016). As the name suggests, this approach allows the explanatory variables in a choice model as latent and stochastic. Doing so helps in recognizing measurement errors in variables (Walker *et al*., 2010), perception errors by individuals (Varotto *et al*., 2017), and even missing data (Sanko *et al*., 2014). A recent study by Varela *et al.* (2018) accommodates measurement errors in travel time and travel cost variables in mode choice models using the *ICLV* approach. In doing so, they examine the magnitude of measurement errors as well as evaluate different distributional assumptions for specifying measurement errors in travel time and travel cost variables. Their analysis suggests larger magnitudes of measurement error in self-reported attributes vis-à-vis that in travel times computed from network-skims. In addition, they highlight that different degrees of measurement error in different variables in the same model could lead to differential biases in the corresponding coefficient estimates and hence a bias in the ratio of coefficient estimates such as willingness-to-pay estimates. In another study, Biswas *et al*. (2019) used the *ICLV* approach to accommodate stochastic travel time variables in route choice models. But all the above studies, while considering stochasticity in variables, maintain deterministic coefficients on those same variables.

In a separate and rather large stream of literature, unobserved taste heterogeneity of individuals (that is, variations in the sensitivity to exogenous variables due to unobserved factors) has been modelled in a multitude of choice contexts. These studies specify random coefficients on alternative attributes through frameworks such as the mixed multinomial logit (Bhat, 2001; Bhat, 2003; Hensher and Greene, 2003; Greene and Hensher, 2003; Batley *et al*., 2004; Hess and Polak, 2005; Mc Fadden and Train, 2000; Revelt and Train, 1998; Brownstone *et al*., 2000; Swait, 2022) and the mixed multinomial probit (Bhat, 2011; Bhat and Sidharthan, 2012; Patil *et al*., 2017; Dubey *et al*., 2022). Regardless of the approach used to accommodate unobserved heterogeneity in response to exogenous variables, this stream of literature does not consider stochasticity in the exogenous variables themselves. In fact, to the best of our knowledge, no study has attempted to recognize and disentangle the two sources of variability – stochasticity in explanatory variables and unobserved heterogeneity in response to those variables. This is because typical mixed logit/probit and *ICLV* model formulations do not allow the simultaneous estimability or identifiability of both sources of variability.

The objective of the current research is to formulate a choice modelling framework that allows the analyst to accommodate stochasticity in explanatory variables and random coefficients on such variables. In addition, the study aims at applying the proposed framework to disentangle travel time variability from unobserved heterogeneity in response to travel time in travel choice models. To this end, we formulate an integrated choice and stochastic variable (*ICSV*) modelling framework with random coefficients (RC) in its choice model.[2] We show that the *ICSV-RC* framework allows the identification of stochasticity in travel time as well as random heterogeneity in response to travel time – due to its ability to bring together two (or more) different data sources such as travel time measurements and traveller choices. In addition, we show that ignoring either source of stochasticity – variability in travel time or heterogeneity in response to travel time – results in models with inferior fit to data and a systematic bias in all parameter estimates. Furthermore, ignoring stochasticity in travel time can potentially lead to underestimation of standard errors. We demonstrate such repercussions of ignoring stochastic explanatory variables using simulation experiments in two distinct choice settings – one involving labelled mode choice alternatives and the other involving unlabelled route choice alternatives. [3]

While our formulation is generic and applicable to any choice context, the empirical application in this study pertains to truck route choice and travel time measurements derived from a large truck-GPS dataset in Florida. Using this empirical data, we demonstrate the applicability of the *ICSV-RC* framework for identifying variability in travel time and a random coefficient on

[2] In the field of choice modelling, the incorporation of latent psychological constructs such as attitudes and perceptions as explanatory variables within the random utility maximization framework assumes the form of a hybrid model that is typically referred to as the Integrated Choice and Latent Variable (*ICLV*) model (for further reading, one may refer to Ben-Akiva *et al*. (2002), Bhat and Dubey (2014), Alvarez-Daziano and Bolduc (2013) and Vij and Walker (2016)). In the *ICLV* framework, stochastic variables are typically used to represent latent psychological constructs such as attitudes and perceptions of the individuals making choices. In this paper, we use the more general label of Integrated Choice and "Stochastic" variable (*ICSV*) framework to recognize that individual latent attitudes/perceptions are but only one form of a stochastic variable within an integrated choice context; the stochasticity in the variable can also derive from inherent variability or measurement error or other sources.

[3] A relevant question in this context is which exogenous variables to consider stochastic. This can be tricky to decide and depends on the empirical context. But some general concepts can be applied. In terms of socio-demographics, it is not unreasonable to think that the sample used in estimation is such that the demographics (across sample observations) represent the population demographics at large. Thus, it may be less inappropriate to believe that we would get effectively similar "fixed" demographic attributes in repeated sampling. However, this is not the case for intra-individual travel time measures, as we discuss in the paper. Thus, reasonable assumptions may be made during model specification to decide on which exogenous variables to consider as stochastic.

travel time. We then compare the *ICSV-RC* model with simpler versions of it – one without random coefficients and one without variability in travel time – to highlight the importance of accounting for both sources of variability.[4]

The rest of this paper is structured as follows. Section 2 discusses the *ICSV-RC* modelling framework in the context of an integrated model of traveller choice and stochastic travel time. Then, Section 2.2 presents a maximum simulated likelihood based simultaneous estimator for the proposed model. In addition, a sequential (two-step) estimator is discussed. Further, simpler versions of the proposed *ICSV-RC* model that ignore either the stochasticity in travel time or unobserved heterogeneity in the coefficient of travel time are derived in Section 2.3. Section 3 discusses the reason for bias in parameter estimates (and the direction of bias) in models that do not incorporate variability in stochastic variables. Section 4 presents simulation experiments for two different travel choice settings, one involving labelled mode choice alternatives and the other involving unlabelled route choice alternatives. The simulations and findings for the mode choice setting are discussed in detail in this section, whereas those for the route choice setting are presented in Appendix A. Section 5 presents the empirical results and findings in the context of truck route choice in Florida, USA. Section 6 concludes the study and identifies directions for future research.

2 MODEL FRAMEWORK

2.1. Model Formulation

The *ICSV-RC* framework is formulated to jointly model the observed travel times for the travel choice alternatives available to the traveller and the traveller's choice[5]. The travel time component of the integrated model helps in characterizing the alternative-specific travel time distributions that

---

[4] The study by Biswas *et al*. (2019) used an ICSV framework to consider route-level travel time as a stochastic variable in a truck route choice model using the same empirical dataset. However, that study, like all other studies that we are aware of that consider stochasticity in exogenous variables, does not consider heterogeneity in travellers' response to travel time. Further, there was no recognition of the importance of (or the discussion of any method) to account for stochasticity in travel time as well as heterogeneity in the sensitivity to travel time. In contrast, the current study focuses on the issue of both travel time and its coefficient being stochastic, the identification of both sources of stochasticity, and the repercussions of ignoring any of the two sources of stochasticity. It is also worth noting that the ICLV model structure in Biswas *et al*. (2019) study is based on a multinomial probit (MNP) structure, which is not easy to use for accommodating both sources of stochasticity. In the current paper, we use the logit-based kernel, which makes it easier to accommodate both sources of stochasticity within the *ICSV-RC* framework.

[5] We refer to route choice of trips and mode choice of an individual by the common term *traveller's choice*. Such terminology is adopted to view the framework as a tool relevant to a broad range of choice settings. Further, in the same vein, the decision-maker is referred to as the *traveller*.

reflect variability in the travel conditions in the network, while also recognizing measurement errors in the travel times observed by the analyst. Simultaneous to the estimation of the travel time distributions, the stochastic travel time variable is used as an explanatory variable in a mixed multinomial logit-based model of traveller choice (route choice or mode choice) with a random coefficient specified on it. Note that the distributional forms of both travel time and its coefficient are assumed to be known *a priori*, but the parameters of those distributions must be estimated.

Here, we present the notational preliminaries for the *ICSV-RC* model. Denote $\mathbf{J}_n$ = $\{1,2,\ldots,i,\ldots,j,\ldots J_n\}$ as the set of all alternatives available to a traveller n (or trip n), where J_n is the total number of alternatives available to the traveller. In a route-choice setting, $\mathbf{J}_n$ represents route alternatives, and, in a mode-choice setting, $\mathbf{J}_n$ represents travel mode alternatives. For each such alternative i, we define a set $\mathbf{M}_{ni}$ of variables for travel time measurements, where $\mathbf{M}_{ni} = \{OTT_{ni1}, OTT_{ni2}, \ldots, OTT_{nim}, \ldots OTT_{niM_i}\}$ and OTT_{nim} represents the m^{th} measurement (or observation) of travel time associated with alternative i for traveller n. Let ott_{nim} denote a realization of OTT_{nim} or an observed value of OTT_{nim} in the data. The number of measurements $|\mathbf{M}_{ni}|$ may vary across choice alternatives, while some alternatives may not have any measurements. Let $\boldsymbol{OTT}_n$ stack the $\mathbf{M}'_{ni}$ vectors of all J_n alternatives into a column vector of size $\left[\sum_{i=1}^{J_n}|\mathbf{M}_{ni}|\right] \times 1$. Let $\mathbf{m}_{ni}$ be the vector of realizations (i.e., ott_{nim}) of $\mathbf{M}_{ni}$ in the data, and let $\boldsymbol{ott}_n$ stack the $\mathbf{m}'_{ni}$ vectors of all J_n alternatives into a column vector. Further, define y_{ni} as an indicator whether alternative i was chosen by the traveller n. y_{ni} assumes the value 1 if the alternative i is chosen and is zero otherwise.

2.1.1 Structural equations for the ICSV-RC model

Define the utility (U_{ni}) associated with choosing an alternative i by a traveller n (or for trip n) as:

$$U_{ni} = \gamma_{n,TT^*} TT^*_{ni} + \boldsymbol{\theta}' \boldsymbol{x}_{ni} + \varepsilon_{ni} \tag{1}$$

In the above equation, TT^*_{ni} is the stochastic travel time variable for alternative i and γ_{n,TT^*} is its coefficient (which is specified as a random parameter). In the current study, we assume $\gamma_{n,TT^*} = \mu_{\gamma_{TT^*}} + \sigma_{\gamma_{TT^*}} z_n$; $z_n \sim N(0,1)$, albeit one can explore several other distributions such as lognormal or truncated normal. Further, $\boldsymbol{x}_{ni}$ is a W×1 vector of other, deterministic (free of measurement errors) attributes of alternative i and $\boldsymbol{\theta}$ is the corresponding vector of coefficients (some or all of which may be assumed as random parameters, as in the case of typical mixed multinomial choice models); and ε_{ni} is a standard Gumbel distributed error term assumed to be independent and identically distributed (IID) across all choice alternatives and travellers.[6]

Note that the stochastic travel time variable (TT^*_{ni}) can be specified using different functional forms depending on the model setup. The simplest approach is to assume that travel time for any trip follows an *a priori* distribution with the same parameters. For example, Walker *et al*. (2010) represent mode-specific travel times in a mode choice model using a normal distribution whose parameters are estimated using available measurements of travel times. A second approach is to specify the mode-specific travel time distribution as a function of mode-specific inverse speed (i.e., the time it takes to travel unit distance) and travel distance. That is, for a mode *i*, the travel time distribution may be expressed as below:

$$TT^*_{ni} = \theta_{ni} d_{ni} \tag{2}$$

In the above equation, d_{ni} denotes the travel distance by mode *i* between the origin and destination of the traveller. θ_{ni} is the corresponding coefficient, which can be interpreted as the inverse speed of mode *i*. Here, θ_{ni} is considered random to accommodate variability in travel conditions.

A third approach, applicable in route choice models, is to represent the travel time distribution using a structural equation that specifies route-level travel time as a function of the underlying route structure, as below:

---

[6] The exogenous variables in $\boldsymbol{x}_{ni}$ include observed route attributes, such as travel costs and tolls, socio-demographic and land-use variables specific to the decision-maker, and the interaction of such variables with observed level-of-service variables. In route choice models, however, since the choice alternatives are typically not labelled, decision-maker variables with alternative specific coefficients and alternative-specific constants are not included. In addition, to recognize physical overlap between different route alternatives available for a trip, it is common to correct the utility functions by including the natural logarithm of a route-specific path size (PS) attribute as an explanatory variable (see Ben-Akiva and Bierlaire (1999) for details on the path size attribute). The PS attribute accommodates correlations between route alternatives due to physical overlap between routes. However, correlations between route alternatives might also arise due to unobserved factors that are not attributable to physical overlap. For example, two routes passing through different sections of a major named highway may share unobserved effects due to unobserved characteristics specific to that named highway, even if the two routes do not overlap. To capture such correlations, one can use error components proposed by Frejinger and Bierlaire (2007), as illustrated below by modifying the utility function in Equation (1) as follows:

$$U_{ni} = \gamma_{n,TT^*} TT^*_{ni} + \boldsymbol{\theta}' \boldsymbol{x}_{ni} + \sigma_a \sqrt{L_{in,a}} \phi_{na} + \sigma_b \sqrt{L_{in,b}} \phi_{nb} + \varepsilon_{ni}.$$

In this expression, $L_{in,a}$, $L_{in,b}$ are the distances covered by route *i* on the roads labelled *a* and *b* respectively. ϕ_{na} and ϕ_{nb} are independent standard normal random variables, assumed to be IID across observations. σ_a and σ_b are parameters to be estimated. Such an error components specification helps capture the perceptual correlations among route alternatives passing through a same labelled road without necessarily overlapping (Frejinger and Bierlaire, 2007). In the rest of the formulation, we do not include these error components for simplicity in notation. But we do include error components in both our simulation experiments and the empirical application for route choice.

$$TT_{ni}^* = \sum_{l=1}^{L} \beta_{nl} d_{nil} + \sum_{q=1}^{Q} \gamma_{nq} r_{niq} \quad (3)$$

Here, d_{nil} denotes the length of roadway links of type *l* on route *i* (length of interstates, length of arterials, length of local roads, etc.) for traveller *n* on their trip, r_{niq} denotes the number of nodes of type *q* on route *i* (no. of left turns, no. of right turns, etc.), and *L* and *Q* denote the total number of roadway link types and stop types, respectively. Further, β_{nl} is the random coefficient on d_{nil}, which may be interpreted as the inverse speed on roadway link of type *l* (i.e., the time it takes to traverse unit length of a roadway of type *l*) on route *i*. γ_{nq} is the random coefficient on r_{niq}, which may be interpreted as the time it takes a vehicle to cross a node of type *q*. The stochasticity in random coefficients $\beta_{nl}(l = 1, \ldots, L)$ and $\gamma_{nq}(q = 1, \ldots, Q)$ helps capture variability in TT_{ni}^* due to variability in travel conditions on different types of links and nodes. In vector form, the structural equation for TT_{ni}^* may be written as:

$$TT_{ni}^* = \boldsymbol{B}_{nl}{}' \boldsymbol{D}_{nil} + \boldsymbol{\Gamma}_{nq}{}' \boldsymbol{R}_{niq} \quad (4)$$

where, $\boldsymbol{D}_{nil} = [\, d_{ni1}, d_{ni2}, \ldots, d_{niL}]'$ is the vector of lengths of roadway links of each of L types on the route and $\boldsymbol{R}_{niq} = [\, r_{ni1}, r_{ni2}, \ldots, r_{niQ}]'$ is the vector of number of nodes of each of Q types. Further, $\boldsymbol{B}_{nl} = [\beta_{n1}, \beta_{n2}, \ldots \beta_{nL}]\,'$ is the vector of random coefficients on $\boldsymbol{D}_{nil}$ and $\boldsymbol{\Gamma}_{nq} = [\gamma_{n1}, \gamma_{n2}, \ldots \gamma_{nQ}]'$ is the vector of random coefficients on $\boldsymbol{R}_{niq}$.

2.1.2 Measurement equations for the ICSV-RC model

Equation (1) for the multinomial choice model involves both the stochastic travel time variable (TT_{ni}^*) and its coefficient (γ_{n,TT^*}), which is also random. This necessitates two separate sources of data to identify each of these distributions through the integrated model. The measurements used to identify taste heterogeneity of travellers (i.e., the distribution of γ_{n,TT^*}) are the observed choices (y_{ni}) for each traveller *n*. Invoking the utility maximization theory, the measurement equation for the observed choices can be written as:

$$y_{ni} = 1, \text{ if } U_{ni} > U_{nj} \forall J_n \in \mathbf{J}_n, j \neq i$$
$$= 0, \text{ otherwise} \quad (5)$$

The measurements used to identify travel time variability (that is, the distribution of the random parameters specified in Equation (2) or Equation (3) for TT^*_{ni}) are the observed travel times (OTT_{nim}) obtained from GPS data or other such data sources. Specifically, the travel time measurements (OTT_{nim}) may be specified as a sum of the stochastic travel time function (TT^*_{ni}) along with an additive noise term to represent measurement errors, as below:

$$OTT_{nim} = TT^*_{ni} + \xi_{nim}, \forall\, m \in \mathbf{M}_{ni} \tag{6}$$

Here, ξ_{nim} is a noise term capturing the measurement error in OTT_{nim} and assumed to be normally distributed, $\xi_{nim} \sim N(0, \rho)$, with variance ρ to be estimated.

As discussed in Biswas *et al*. (2019), the stochastic travel time function in Equation (2), identified due to available travel time measurements for some observations in the data, can be simultaneously used to *impute* the travel time distribution for observations without travel time measurements. Doing so helps in utilizing partial measurement data, where travel time measurements may not be available for all observations or choice alternatives, for estimating the integrated model.

2.2. Model System Estimation

Equation (1) along with the equation for the travel time function (Equation (2) or (4), depending on the functional form of the stochastic variable under consideration), and Equations (5) and (6) are brought together into an *ICSV-RC* framework for deriving the joint likelihood of travel time measurements and traveller choices in the observed data. Furthermore, distributional assumptions are made on the stochastic components of the formulation to derive the likelihood function for estimating model parameters.

In the *ICSV-RC* model, let $\mathbf{\Theta} = \{\mu_{\gamma_{TT^*}}, \sigma_{\gamma_{TT^*}}, \boldsymbol{\theta}, Vech(\boldsymbol{B}_{nl}), Vech(\boldsymbol{\Gamma}_{nq}), \rho\}$ denote the full set of parameters to be estimated in the integrated model system, where $Vech\,(.)$ is an operator used to represent the vector of the parameters inside the parentheses. For later use, define $\widetilde{\mathbf{\Theta}} = \{\mu_{\gamma_{TT^*}}, \sigma_{\gamma_{TT^*}}, \boldsymbol{\theta}\}$; $\breve{\mathbf{\Theta}} = \{Vech(\boldsymbol{B}_{nl}), Vech(\boldsymbol{\Gamma}_{nq})\}$; and $\overline{\mathbf{\Theta}} = \{\mu_{\gamma_{TT^*}}, \sigma_{\gamma_{TT^*}}\}$. Let $\boldsymbol{X}_n$ = $[\boldsymbol{x}_{n1}, \boldsymbol{x}_{n2}, \ldots, \boldsymbol{x}_{nJ_n}]'$; $\boldsymbol{D}_n$ = $[(\boldsymbol{D}_{n1}', \boldsymbol{D}_{n2}', \ldots, \boldsymbol{D}_{nJ_n}'), (\boldsymbol{R}_{n1}', \boldsymbol{R}_{n2}', \ldots, \boldsymbol{R}_{nJ_n}')]'$, and let $\boldsymbol{TT}^*_n$= $[TT^*_{n1}, TT^*_{n2}, \ldots, TT^*_{nJ_n}]'$ be the J_n×1 vector of stochastic travel times for all alternatives in the choice set. Next, to denote the probability density of the stochastic components, let $f(.)$ be the PDF of OTT_{njm} given TT^*_{nj}, let $g(.)$ be the PDF of TT^*_{nj}, and let $h(.)$ be the PDF of γ_{n,TT^*}. Using this notation, we discuss the following two approaches to estimate the model parameters in this section: (1) simultaneous estimation and (2) sequential estimation.

2.2.1 Simultaneous estimation

To simultaneously estimate all the parameters of the proposed *ICSV-RC* model, the joint likelihood for the i^{th} alternative being chosen by traveller *n* along with the available observed measurements of route-level travel time is given by:

$$\mathcal{L}\,(y_{ni}=1,\boldsymbol{OTT}_n=\boldsymbol{ott}_n|\,\boldsymbol{D}_n,\boldsymbol{X}_n\,,\boldsymbol{\Theta})$$

$$=\int\limits_{\gamma_{n,TT^*}}\int\limits_{\boldsymbol{TT}_n^*}\mathcal{L}\left(y_{ni}=1,\boldsymbol{OTT}_n=\boldsymbol{ott}_n\,\middle|\,\boldsymbol{TT}_n^*,\boldsymbol{X}_n\,,\widetilde{\boldsymbol{\Theta}},\rho\right)\prod_{j=1}^{J_n}g\left(TT_{nj}^*\middle|\boldsymbol{D}_n,\breve{\boldsymbol{\Theta}}\right)h\left(\gamma_{n,TT^*}\middle|\overline{\boldsymbol{\Theta}}\right)d(\boldsymbol{TT}_n^*)d(\gamma_{n,TT^*}) \tag{7}$$

$$=\int\limits_{\gamma_{n,TT^*}}\int\limits_{\boldsymbol{TT}_n^*}\mathcal{L}\left(y_{ni}=1\,\middle|\boldsymbol{OTT}_n=\boldsymbol{ott}_n,\boldsymbol{TT}_n^*,\boldsymbol{X}_n\,,\widetilde{\boldsymbol{\Theta}}\right)$$
$$\times\prod_{j=1}^{J_n}\prod_{m=1}^{M_{ni}}f\left(ott_{njm}\middle|TT_{nj}^*,\rho\right)\prod_{j=1}^{J_n}g\left(TT_{nj}^*\middle|\boldsymbol{D}_n,\breve{\boldsymbol{\Theta}}\right)h\left(\gamma_{n,TT^*}\middle|\overline{\boldsymbol{\Theta}}\right)d(\boldsymbol{TT}_n^*)d(\gamma_{n,TT^*}) \tag{8}$$

Note that the likelihood $\mathcal{L}\left(y_{ni}=1\,\middle|\,\boldsymbol{OTT}_n=\boldsymbol{ott}_n,\boldsymbol{TT}_n^*,\boldsymbol{X}_n\,,\widetilde{\boldsymbol{\Theta}}\right)$ in the right side of Equation (8) can simply be written as $\mathcal{L}\left(y_{ni}=1\,\middle|\,\boldsymbol{TT}_n^*,\boldsymbol{X}_n\,,\widetilde{\boldsymbol{\Theta}}\right)$. This is because conditional on the actual distribution ($\boldsymbol{TT}_n^*$) of the input variable, the analyst's measurement ($\boldsymbol{OTT}_n$) of that variable does not matter to y_{ni}. That is, the choice of a route conditional on the actual route-level travel times does not depend on the measured travel times. Therefore, the joint likelihood may be written as:

$$\mathcal{L}\,(y_{ni}=1,\boldsymbol{OTT}_n=\boldsymbol{ott}_n|\,\boldsymbol{D}_n,\boldsymbol{X}_n\,,\boldsymbol{\Theta})$$

$$=\int\limits_{\gamma_{n,TT^*}}\int\limits_{\boldsymbol{TT}_n^*}\mathcal{L}\left(y_{ni}=1\,\middle|\,\boldsymbol{TT}_n^*,\boldsymbol{X}_n\,,\widetilde{\boldsymbol{\Theta}}\right)$$
$$\times\prod_{j=1}^{J_n}\prod_{m=1}^{M_{ni}}f\left(ott_{njm}\middle|TT_{nj}^*,\rho\right)\prod_{j=1}^{J_n}g\left(TT_{nj}^*\middle|\boldsymbol{D}_n,\breve{\boldsymbol{\Theta}}\right)h\left(\gamma_{n,TT^*}\middle|\overline{\boldsymbol{\Theta}}\right)d(\boldsymbol{TT}_n^*)d(\gamma_{n,TT^*}) \tag{9}$$

Replacing $\mathcal{L}\left(y_{ni}=1\,\middle|\boldsymbol{TT}_n^*,\boldsymbol{X}_n\,,\widetilde{\boldsymbol{\Theta}}\right)$ in the above expression with the logit choice probability expression, the joint likelihood may be written as:

$$\mathcal{L}\,(y_{ni}=1,\boldsymbol{OTT}_n=\boldsymbol{ott}_n|\,\boldsymbol{D}_n,\boldsymbol{X}_n\,,\boldsymbol{\Theta})$$

$$= \int_{\gamma_{n,TT^*}} \int_{\boldsymbol{TT}_n^*} \frac{\exp\left(\gamma_{n,TT^*}TT_{ni}^* + \boldsymbol{\theta}'\boldsymbol{x}_{ni}\right)}{\sum_{j=1}^{J_n}\exp\left(\gamma_{n,TT^*}TT_{nj}^* + \boldsymbol{\theta}'\boldsymbol{x}_{nj}\right)} \times \prod_{j=1}^{J_n}\prod_{m=1}^{M_{ni}} f\left(ott_{njm}\middle|TT_{nj}^*,\rho\right)\prod_{j=1}^{J_n} g\left(TT_{nj}^*\middle|\boldsymbol{D}_n,\breve{\boldsymbol{\Theta}}\right) h\left(\gamma_{n,TT^*}\middle|\overline{\boldsymbol{\Theta}}\right) d(\boldsymbol{TT}_n^*)\, d(\gamma_{n,TT^*}) \tag{10}$$

The dimensionality of integration in the above equation is the total number of unique random parameters in U_{ni} (except the IID Gumbel error terms). This includes the number of random coefficients in TT_{ni}^*, one random coefficient for γ_{n,TT^*}, etc. One can use the maximum simulated likelihood method to optimize the above likelihood function and estimate the parameters of the *ICSV-RC* model.

For observations without any measurements of travel time, Equation (10) becomes:

$$\mathcal{L}\,(y_{ni}=1|\,\boldsymbol{D}_n,\boldsymbol{X}_n\,,\boldsymbol{\Theta})$$

$$= \int_{\gamma_{n,TT^*}} \int_{\boldsymbol{TT}_n^*} \frac{\exp\left(\gamma_{n,TT^*}TT_{ni}^* + \boldsymbol{\theta}'\boldsymbol{x}_{ni}\right)}{\sum_{j=1}^{J_n}\exp\left(\gamma_{n,TT^*}TT_{nj}^* + \boldsymbol{\theta}'\boldsymbol{x}_{nj}\right)} \prod_{j=1}^{J_n} g\left(TT_{nj}^*\middle|\boldsymbol{D}_n,\breve{\boldsymbol{\Theta}}\right) h\left(\gamma_{n,TT^*}\middle|\overline{\boldsymbol{\Theta}}\right) d(\boldsymbol{TT}_n^*)\, d(\gamma_{n,TT^*}) \tag{11}$$

In this context, the travel time measurement data of other observations helps estimate the parameters describing the distribution of $\boldsymbol{TT}_n^*$. These parameters are, in turn, used simultaneously in Equation (11) to inform or *impute* the stochastic travel time function for observations without travel time measurements. The same happens in Equation (10) as well, for observations with travel time measurements for some alternatives and no measurements for other alternatives.

2.2.2 Sequential estimation (two-step estimation)

One can estimate the parameters of the model system in a sequential, two-step procedure. In this approach, the first step involves the estimation of the parameters ($\breve{\boldsymbol{\Theta}}$) of $\boldsymbol{TT}_n^*$ and ρ using Equation (6), which is a regression model with random parameters. The second step involves a mixed logit model with stochastic explanatory variables (*ML-SV* model) that utilizes the *a priori* known distribution of $\boldsymbol{TT}_n^*$ as stochastic explanatory variables and specifies the coefficient (γ_{n,TT^*}) on $\boldsymbol{TT}_n^*$ as random. The likelihood expression of such an *ML-SV* model with stochastic travel time variables and a random coefficient on the stochastic variables is:

$$\mathcal{L}\left(y_{ni}=1 \mid \boldsymbol{D}_n, \boldsymbol{X}_n, \boldsymbol{\Theta}\right)$$

$$= \int_{\gamma_{n,TT^*}} \int_{\boldsymbol{TT}_n^*} \frac{\exp\left(\gamma_{n,TT^*} TT_{ni}^* + \boldsymbol{\theta}' \boldsymbol{x}_{ni}\right)}{\sum_{j=1}^{J_n} \exp\left(\gamma_{n,TT^*} TT_{nj}^* + \boldsymbol{\theta}' \boldsymbol{x}_{nj}\right)} \prod_{j=1}^{J_n} g\left(TT_{nj}^* \middle| \boldsymbol{D}_n, \breve{\boldsymbol{\Theta}}\right) h\left(\gamma_{n,TT^*} \middle| \overline{\boldsymbol{\Theta}}\right) d(\boldsymbol{TT}_n^*) d(\gamma_{n,TT^*}) \quad (12)$$

The parameters to be estimated in the above *ML-SV* model are those in $\boldsymbol{\theta}$ and $\overline{\boldsymbol{\Theta}}$, since the estimates of parameters in $\breve{\boldsymbol{\Theta}}$ and ρ are available from the first step estimation.

It is worth noting here that the simultaneously estimated *ICSV-RC* model in Section 2.2.1 and the sequentially estimated *ML-SV* model presented here, both account for the two sources of stochasticity discussed earlier – stochastic travel times and random coefficient on stochastic travel times. The difference between the two models is in the estimation of parameters. The former involves a simultaneous estimation of all parameters (through maximizing the joint likelihood of observed travel choices and observed travel times). The latter involves a two-step estimation. Both these approaches are associated with pros and cons, as discussed next.

Between the simultaneous and sequential estimation approaches, the simultaneous estimation approach is preferable for the following reasons. (1) Simultaneous estimation results in more efficient estimates when compared to sequential estimation approaches (Heckman, 1976; Heckman, 1979). In this context, the two-step approach might require adjustment of standard errors of its parameter estimates, which can potentially involve cumbersome procedures. (2) The two-step estimation can potentially result in biased estimates when there is endogeneity between $\boldsymbol{TT}_n^*$ and the travel choice (y_{ni}) when common unobserved factors influence both $\boldsymbol{TT}_n^*$ and travel choice or when there is simultaneity between $\boldsymbol{TT}_n^*$ and travel choice. In this context, the simultaneous estimation provides scope to address such endogeneity by more easily allowing for correlations between $\boldsymbol{TT}_n^*$ and the utility functions of the choice model component.

On the other hand, the sequential estimation approach might be preferred for the following reasons. (1) It may be computationally easier to estimate the parameters in two different steps, as opposed to a joint estimation of all parameters in a single step. (2) In the first step of the two-step approach, sophisticated modelling techniques such as semi non-parametric approaches (Fosgerau and Fukuda, 2012; Rahmani *et al*., 2015), data-driven deep learning methods (Zhang *et al*., 2019; James, 2021), etc., can be used to describe the distribution of $\boldsymbol{TT}_n^*$. Doing so can potentially help characterize $\boldsymbol{TT}_n^*$ in a better way than simple, parametric forms for $\boldsymbol{TT}_n^*$ that are easier to embed in joint estimation. Implementing such sophisticated approaches to characterize $\boldsymbol{TT}_n^*$ within a joint estimation framework can potentially be cumbersome.

In the current research, given our focus is on the importance of accommodating the two sources of stochasticity – stochasticity in alternative attributes and random coefficients on the stochastic attributes – using appropriate data sources, we employ the simultaneous estimation approach with a simple parametric specification of $\boldsymbol{TT}_n^*$. Exploration of more advanced approaches (such as semi non-parametric and data-driven methods) to characterize the distribution of $\boldsymbol{TT}_n^*$, a detailed evaluation of the pros and cons of doing so with simultaneous and sequential estimation approaches, and the correction of standard errors in the sequential estimation approach needs to be addressed in a future extension of this work.

2.3. Alternative Model Structures

In this section, we discuss alternative model structures that are simpler versions of the above discussed *ICSV-RC* model.

2.3.1 Integrated model with stochastic travel time and fixed coefficient on travel time (ICSV model)

A restricted form of the *ICSV-RC* model is obtained by specifying the coefficient γ_{n,TT^*} as fixed (denoted by $\bar{\gamma}_{TT^*}$ in the current model) instead of random. For such an *ICSV* model with a fixed coefficient on travel time, denote the full set of parameters to be estimated as $\boldsymbol{\Psi} = \{\bar{\gamma}_{TT^*}, \boldsymbol{\theta}, \widetilde{\boldsymbol{\Theta}}, \rho\}$, where $\bar{\gamma}_{TT^*}$ is the deterministic coefficient on travel time, and $\breve{\boldsymbol{\Theta}}$ is $\{Vech(\boldsymbol{B}_{nl}), Vech(\boldsymbol{\Gamma}_{nq})\}$ as defined earlier. This *ICSV* model's joint likelihood expression reduces from the *ICSV-RC* model likelihood in Equation (10) to the following:

$$\mathcal{L}\,(y_{ni} = 1, \boldsymbol{OTT}_n = \boldsymbol{ott}_n |\, \boldsymbol{D}_n, \boldsymbol{X}_n\,, \boldsymbol{\Psi})$$

$$= \int_{\boldsymbol{TT}_n^*} \frac{\exp(\bar{\gamma}_{TT^*} TT_{ni}^* + \boldsymbol{\theta}'\boldsymbol{x}_{ni})}{\sum_{j=1}^{J_n} \exp\left(\bar{\gamma}_{TT^*} TT_{nj}^* + \boldsymbol{\theta}'\boldsymbol{x}_{nj}\right)} \prod_{j=1}^{J_n} \prod_{m=1}^{M_{ni}} f\left(ott_{njm} \middle| TT_{nj}^*, \rho\right) \prod_{j=1}^{J_n} g\left(TT_{nj}^* \middle| \boldsymbol{D}_n, \breve{\boldsymbol{\Theta}}\right) d(\boldsymbol{TT}_n^*) \qquad (13)$$

2.3.2 Mixed logit model with expected travel time and random coefficient on travel time (ML-RC)

This model (*ML-RC*) involves using the expected travel time, $E(TT_{ni}^*)$, obtained using the parameter estimates of the stochastic travel time ($\boldsymbol{TT}_n^*$) function (instead of using the entire stochastic distribution for travel time) as an explanatory variable in the choice utility function. As discussed in the case of *ML-SV* model, the parameters of $\boldsymbol{TT}_n^*$ are estimated in an *a priori* step.

The coefficient on $E(TT_{ni}^{*})$, denoted as $\gamma_{n,TT^{*}}$, is specified to be random to capture travellers' taste heterogeneity. This specification leads to a standard mixed logit model that is long established in the existing literature, with the following likelihood expression:

$$\mathcal{L}\left(y_{ni}=1\middle|\, E(TT_{ni}^{*}), \boldsymbol{X}_n, \boldsymbol{\theta}, \overline{\boldsymbol{\Theta}}\right)$$

$$= \int_{\gamma_{n,TT^{*}}} \frac{\exp\left(\gamma_{n,TT^{*}} E(TT_{ni}^{*}) + \boldsymbol{\theta}'\boldsymbol{x}_{ni}\right)}{\sum_{j=1}^{J_n} \exp\left(\gamma_{n,TT^{*}} E(TT_{nj}^{*}) + \boldsymbol{\theta}'\boldsymbol{x}_{nj}\right)} h\left(\gamma_{n,TT^{*}}\middle|\overline{\boldsymbol{\Theta}}\right) d(\gamma_{n,TT^{*}}) \tag{14}$$

2.3.3 Multinomial logit (MNL) and mixed logit with error components (ML-EC)

Simplifying the above model further by treating the travel time coefficient as deterministic results in the multinomial logit (*MNL*) model with expected trave time ($E(TT_{ni}^{*})$) as one of the variables in the utility function. If the model specification includes error components to account for correlations among choice alternatives, then it would become a simple mixed logit model with error components (*ML-EC*) that does not accommodate stochasticity in travel time nor unobserved heterogeneity in the coefficient on travel time.

3 ESTIMATION BIAS DUE TO IGNORING STOCHASTICITY IN VARIABLES

In models that ignore stochasticity in explanatory variables (for example, travel time), the parameter estimates for the coefficients of such variables as well as those of other variables demonstrate a bias. Here, we discuss the nature of the bias. To do so, we consider two possible types of stochasticity in explanatory variables – additive stochasticity and multiplicative stochasticity – and consider each of the following cases for the coefficients on such variables: (1) the coefficient on the stochastic variable is random and follows a distribution that allows additive separability of the location parameter and the scale parameter (e.g., normal distribution); (2) the coefficient on the stochastic variable is random and follows a distribution that does not allow additive separability of the location parameter and the scale parameter (e.g., lognormal distribution).

Consider a true model with the following utility function associated with alternative *i* (Note: In this discussion, the subscript *n* for traveller is suppressed for simplicity):

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + \beta X_i^{*} + \varepsilon_i \tag{15}$$

In the above model, $\boldsymbol{Z}_i$ is a vector of deterministic variables with its coefficient vector $\boldsymbol{\theta}$, X_i^* is a stochastic variable with a random coefficient β, and ε_i is an idiosyncratic random error term, assumed to be independent and identically distributed across individuals and alternatives. In addition, assume that β and X_i^* are independent of each other and that $\boldsymbol{Z}_i$ and X_i^* do not offer any information on ε_i (*i.e.*, $E(\varepsilon_i|\boldsymbol{Z}_i) = E(\varepsilon_i|X_i^*) = 0$). Further, let $Var(\varepsilon_i) = \sigma_\varepsilon^2$ and $E(\varepsilon_i) = 0$.

Now, let the measurement available with the analyst for X_i^* be X_i. And let the gap between X_i^* and X_i be represented using an additive error term ν_i; i.e., $X_i^* = X_i + \nu_i$, where $E(\nu_i) = 0$ and $Var(\nu_i) = \sigma_\nu^2$. With an additive error in representing the explanatory variable X_i^*, the utility function in Equation (15) may be rewritten as:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + \beta X_i + \beta \nu_i + \varepsilon_i \tag{16}$$

If we estimate the above true model that recognizes the error ν_i in X_i^*, then the kernel error term for such a model would be the same as ε_i with a variance σ_ε^2. On the other hand, if we ignore the additive error ν_i in X_i^*, then ν_i would get lumped into the kernel error, resulting in the following model with a new kernel error term δ_i:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + \beta X_i + \delta_i \tag{17}$$

In the above model, the new kernel error term δ_i can be expressed as $\beta\nu_i + \varepsilon_i$. The variance of the new kernel error term δ_i in the above model is $Var(\beta\nu_i) + \sigma_\varepsilon^2$, since ε_i and ν_i are independent (because $E(\varepsilon_i|X_i^*) = 0$). That is, the variance of δ_i is greater than the variance (σ_ε^2) of the kernel error term in the true model of Equation (15).

Alternatively, consider a multiplicative specification for the error in X_i^*. That is $X_i^* = X_i\eta_i$, where $E(\eta_i) = 1$ and $Var(\eta_i) = \sigma_\eta^2$. Then, Equation (15) may be re-written as:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + \beta X_i \eta_i + \varepsilon_i \tag{18}$$

If we do not explicitly recognize the error η_i in the above model and include only βX_i in the utility function, it would result in the following model with a new kernel error term ξ_i:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + \beta X_i + \xi_i \tag{19}$$

In the above model, the new kernel error term ξ_i can be expressed as $\beta X_i \eta_i - \beta X_i + \varepsilon_i$. The variance of this term is $Var(\beta X_i \eta_i - \beta X_i) + \sigma_\varepsilon^2$, which is greater than the variance (σ_ε^2) of the kernel error term in the true model of Equation (15).

In summary, regardless of whether the stochasticity in explanatory variables is additive or multiplicative, ignoring such stochasticity results in a model with a kernel error term that has a greater variance than that of the kernel error term in the true model. This result will have a bearing on the direction of bias in the parameter estimates of the model that ignores stochasticity in X_i^*. Since the variance of the kernel error δ_i in Equation (17) and that of the kernel error ξ_i in Equation (19) are greater than that of the true model in Equation (15), one can expect the estimates of the coefficient vector $\boldsymbol{\theta}$ to be biased toward zero.[7] This is because, regardless of the distributional assumption on ε_i (whether it is multivariate normal or Gumbel), the parameter estimates of the utility function are confounded with the scale of the kernel error term.

The nature of bias in the parameter estimates of the distribution for β depends on its distributional assumption. If β follows a distribution that allows its location parameter to be additively separable from its scale parameter, such as $\beta \sim N(\mu_\beta, \sigma_\beta^2)$ or $\beta = \mu_\beta + \sigma_\beta z_\beta$, where z_β is a standard normal variate, then the estimates of both μ_β and σ_β will be biased toward zero. This is because the utility function of Equation (17) may be written as:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + (\mu_\beta + \sigma_\beta z_\beta)X_i + \delta_i \tag{20}$$

and the utility function of Equation (19) may be written as:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + (\mu_\beta + \sigma_\beta z_\beta)X_i + \xi_i \tag{21}$$

It is clear from the above two utility functions that the estimates of both μ_β and σ_β will be biased toward zero, because the variances of the kernel errors δ_i and ξ_i are greater than the variance of the true model's kernel error term.

[7] Note that this trend in bias (toward zero) is similar to the bias one can expect when a normally distributed random coefficient in the utility function is incorrectly specified as deterministic, which is an established result in the literature (Brownstone *et al*., 2000; Cherchi and Ortúzar, 2008; Swait and Bernardino, 2000, Train, 1998).

On the other hand, if β follows a lognormal distribution such that $\beta = \exp\,(\mu_\beta + \sigma_\beta z_\beta)$, where z_β is a standard normal variate, the utility function of Equation (17) may be written as:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + exp(\mu_\beta) \times \exp(\sigma_\beta z_\beta) \times X_i + \delta_i \tag{22}$$

and the utility function of Equation (19) may be written as:

$$U_i = \boldsymbol{\theta}'\boldsymbol{Z}_i + exp(\mu_\beta) \times \exp(\sigma_\beta z_\beta) \times X_i + \xi_i \tag{23}$$

In the above two equations, since μ_β and σ_β are not additively separated, although the variances of the kernel error terms are inflated (when compared to that of the true model), one cannot say with certainty that both μ_β and σ_β will be biased toward zero. Only one of them or both of them might be biased toward zero. It is also possible that one of the two parameter estimates gets biased toward zero while the other gets biased away from zero.

4 SIMULATION EXPERIMENTS

Three distinct sets of simulation experiments were conducted to evaluate parameter recovery of the proposed models and examine the repercussions of ignoring stochasticity in alternative attributes. The first set (Set I) of experiments is for a travel mode choice context with labelled alternatives. In this set of simulations, a truncated normal distribution was assumed for the travel time of one of the modes and a normal distribution was assumed for the coefficient on travel time. The experimental design and the findings from these experiments are presented in Section 4.1.

The second set (Set II) of experiments is also for a mode choice setting. However, in these experiments, the travel times of one of the modes and the travel time coefficient were assumed to follow lognormal distributions. The experimental design and the findings from these experiments are presented in Section 4.2.

The third set (Set III) of experiments is for a route choice context with unlabelled alternatives. In these simulations, normal distributions were assumed for both route-level travel times and the coefficient on travel times. The experimental design and the findings from these experiments are presented in Appendix A.

In addition to the above three sets of simulation experiments, we conducted an additional fourth set (Set IV) of experiments, for the mode choice setting, to evaluate the effect of incorrect distributional assumptions on mode-specific travel times and the corresponding coefficient. Also, we conducted a fifth set (Set V) of experiments for the mode choice setting to evaluate the influence of treating travel time as stochastic when there are no measurement errors in travel times (i.e., travel times are not stochastic). The experimental design and the findings from these two sets of experiments are presented in Appendices B and C, respectively.

The travel mode choice context – common to all but the third set of experiments – involves three labelled alternatives, for which data were generated to reflect travel conditions akin to those in Bengaluru, India. The third set of simulation experiments is that of route choice with unlabelled alternatives, for which synthetic data were generated to mimic the empirical dataset we used for the empirical analysis on route choice from Florida, USA.

4.1 Simulation experiment Set I: Mode choice setting with truncated normal distribution assumption for bus travel time and normal coefficient on travel time

4.1.1 Simulation design

The mode choice simulation experiments for set I were conducted using 200 simulated datasets, each dataset with a sample size of 5,000 individuals. Three modes – bus, car, and walk – were considered. Of these, bus and car were assumed to be available for all individuals, while the walk mode was assumed to be available for travel distances of 10 km or less. Two mode-specific attributes were considered in the utility functions: travel time and travel cost, as shown below:

$$U_b = \beta_{0b} + \gamma_{TT^*} TT_b{}^* + \beta_C TC_b + \varepsilon_b \quad (24)$$

$$U_c = \gamma_{TT^*} TT_c + \beta_C TC_c + \varepsilon_c \quad (25)$$

$$U_w = \beta_{0w} + \gamma_{TT^*} TT_w + \varepsilon_w \quad (26)$$

In the above utility functions, travel times of the bus mode ($TT_b{}^*$) were considered stochastic (the specific value is unknown to the analyst) while those of other modes (TT_c and TT_w) were considered to be known to the analyst. To simulate travel times for the bus mode, the equation $TT_b{}^* = \theta_b d_b$ was used, where θ_b is the inverse speed for the bus mode and d_b is the trip distance on bus. θ_b was assumed to follow a left-truncated normal distribution, whose underlying normal distribution mean is 1.50 min/km (which corresponds to a maximum speed of 40 kmph) and standard deviation (SD) is 0.15. The left-truncation value for θ_b is 1.33 min/km (i.e., a maximum bus speed of 45 kmph in the city), which was assumed to be known while the afore-mentioned mean and standard deviation are parameters to be estimated. Trip distance (d_b) is an exogenous variable representing distances travelled in Bengaluru. TC_b and TC_c are travel costs of the bus and car modes, respectively, which were assumed to be free of measurement errors. Data for TC_b were generated to reflect distance-based bus ticket prices in the city, and data for TC_c were generated based on fuel price in Bengaluru and average mileage of a hatchback car model. The travel time coefficient (γ_{TT^*}) was assumed to follow a normal distribution with mean -1.00 and standard deviation (SD) 0.19. The travel cost coefficient (β_C) was assumed to be -0.25. Finally, the error terms (ε_b, ε_c, ε_w) were assumed to be IID standard Gumbel distributed.

Using the above utility functions and the utility maximization principle, 200 mode choice datasets, each comprising 5,000 trips were simulated. For each of the 5,000 trips from each of the 200 datasets, a single measurement of bus travel time (i.e., observed travel time or OTT_{im}) was simulated by adding a normal distributed measurement error to the simulated value of ${TT_b}^*$. The standard deviation of this measurement error was assumed to be 0.95.

4.1.2 Evaluation and discussion

The above-discussed simulated data of mode choices and observed travel times (y_i, OTT_{im}) along with the simulated exogenous variables ($d_b, TC_b, TC_c, TT_c, TT_w$) were used to estimate the models discussed in Section 2. Parameter recovery across the 200 simulated datasets was examined using the metrics summarized below:

(1) *Absolute Percentage Bias* (APB): For a given parameter in the model, APB is the absolute value of the difference between the true parameter value and the mean of the parameter estimates across the 200 simulated datasets – expressed as a percentage of the true parameter value.

(2) *Asymptotic Standard Error* (ASE): ASE for a given parameter is the mean (across the 200 simulated datasets) of the standard errors of the parameter's estimated values.

(3) *Finite Sample Standard Error* (FSSE): FSSE for a given parameter is the standard deviation of the parameter's estimated values across the 200 datasets.

The above set of metrics are used to evaluate the simulation results for this set as well as in the following sub-sections for the other sets of simulation experiments.

Table 1 presents the above evaluation metrics for different models estimated in this study. The true parameter values used for simulating the data are shown in the second column of the table. The next set of columns, under the title "*ICSV-RC model parameter estimates*", shows the parameter recovery metrics for the *ICSV-RC* model. As can be observed from these columns, the parameters of both the travel time model and the mode choice model are recovered accurately (i.e., with low APB) and precisely (i.e., with low standard errors). In addition, the closeness of the FSSE and ASE values suggests that the estimator of the standard error serves as a good approximation to the finite sample efficiency for the sample size considered in the study. Further, it is worth noting from Table 1 that the APB values of the *ICSV-RC* model are lower than those of all other models – *ICSV*, *ML-RC*, *ML-SV* and *MNL* models.

Recall that the *ML-SV* model, similar to the *ICSV-RC* model, also accommodates the two sources of stochasticity, albeit through a sequential estimation approach. Between the choice components of the *ML-SV* and *ICSV-RC* models, the *ML-SV* demonstrates relatively higher APB values than the *ICSV-RC* model. Further, as expected, the *ML-SV* model is associated with higher FSSE values and a wider gap between ASE and FSSE values than the *ICSV-RC* model. At the same time, the *ML-SV* model, although associated with a higher estimation bias and lower efficiency than the *ICSV-RC* model, performs superior to the other models that ignore stochasticity in travel time or in its coefficient. This finding again highlights that there is value in incorporating stochasticity in explanatory variables, even if through a sequential approach.

Next, we turn to the direction of bias in the parameter estimates for the *ML-RC* model that ignores travel time variability when compared to the *ICSV-RC* model that incorporates travel time variability. In Table 1, the column titled "*t-stat. for* H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ML-RC}$" presents the t-test statistics for the null hypothesis that the magnitude of the parameter estimates from the *ICSV-RC* model are statistically the same as those from the *ML-RC* model.[8] As can be observed from the parameter estimates of the two models (*ICSV-RC* and *ML-RC*) and the t-test statistics, both the

[8] The paired t-statistic values for the difference between the parameter estimates of the *ICSV-RC* model (label it Model 1) and those of another model (label it Model 2) were computed using the expression below:

$$t = \left(\frac{\text{Mean}(\widehat{\beta_{m1}}) - \text{Mean}(\widehat{\beta_{m2}})}{\sqrt{ASE_{\widehat{\beta_{m1}}}{}^2 + ASE_{\widehat{\beta_{m2}}}{}^2} - 2 \times Cov(\widehat{\beta_{m1}}, \widehat{\beta_{m2}})} \right)$$

In the above expression, Mean($\widehat{\beta_{m1}}$) and Mean($\widehat{\beta_{m2}}$) are the mean values (across the 200 simulated datasets) of the parameter estimates of the coefficients from Model 1 and Model 2, respectively. $ASE_{\widehat{\beta_{m1}}}$ and $ASE_{\widehat{\beta_{m2}}}$ are the corresponding asymptotic standard errors computed as the averages of the standard errors across the 200 datasets. $Cov(\widehat{\beta_{m1}}, \widehat{\beta_{m2}})$ is the covariance between the parameter estimates across the 200 datasets. The denominator of the above expression represents the standard error of the difference.

estimated mean and standard deviation of the coefficient on travel time are biased towards zero in the *ML-RC* model. Further, the other parameters in the mode choice utility functions also demonstrate a similar trend – a bias towards zero – in the *ML-RC* model when compared to models that incorporate both travel time variability and random coefficient on travel time. These results are in line with the discussion in Section 3 on the repercussions of ignoring stochasticity in explanatory variables. In addition, the parameter estimates in the mode choice model component of the *ICSV* model are also biased towards zero when compared to those in the choice component of the *ICSV-RC* model. Although not shown in the table, similar t-tests suggest rejecting the null hypothesis that the parameter estimates of the *ML-SV* model (which incorporates stochasticity in travel time using sequential estimation) are same as those from the *ML-RC* model. The bias increases further in the *MNL* model that ignores both sources of variability – stochastic variables and random coefficients on those variables. These results highlight the importance of incorporating both sources of variability, ignoring either of which would result in parameter estimates with a systematic bias toward zero.

Furthermore, although not reported in the table, in most of the 200 datasets, the data fit of the mode choice component of the *ICSV-RC* model was statistically superior (as tested by the log-likelihood ratio test) to that of other models that ignore either the randomness in travel time (*ML-RC* model), or randomness in the coefficient of travel time (*ICSV* model), or both sources of stochasticity (*MNL*). The average likelihood values and other metrics such as Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were also much better (by at least a few hundred points) for the mode choice component of the *ICSV-RC* than those of *ICSV*, *ML-RC*, and *MNL* models. These results highlight the importance of incorporating stochasticity in explanatory variables and the coefficients on such variables. Ignoring any of these two sources of variability, when present, can potentially lead to inferior parameter recovery and model fit.

Table 1 Simulation evaluation results for the mode choice setting (Set I): Truncated normal distribution for travel time and normal distributed coefficient on travel time

	True value	***ICSV-RC* model parameter estimates**				***ICSV* model parameter estimates**					***ML-RC* model parameter estimates**					***ML-SV* model parameter estimates**				***MNL* model parameter estimates**			
		Mean	APB	FSSE	ASE	Mean	APB	FSSE	ASE	t-stat. for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ICSV}$	Mean	APB	FSSE	ASE	t-stat. for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ML}$	Mean	APB	FSSE	ASE	Mean	APB	FSSE	ASE
Bus travel time model																							
Inverse speed θ_b: Location parameter	1.50	1.49	0.7	0.002	0.004	1.47	1.8	0.002	0.004	4.07	--	--	--	--	--	1.51	0.7	0.006	0.008	--	--	--	--
Inverse speed θ_b: Scale parameter	0.15	0.14	9.3	0.002	0.003	0.13	14.2	0.002	0.003	1.20	--	--	--	--	--	0.11	26.7	0.005	0.004	--	--	--	--
SD of measurement error	0.95	0.93	2.1	0.045	0.043	0.94	1.3	0.045	0.043	0.16	--	--	--	--	--	1.45	52.6	0.063	0.089	--	--	--	--
Mean APB, FSSE, ASE	--	--	4.1	0.016	0.017	--	5.7	0.017	0.017	--	--	--	--	--	--	--	26.7	0.025	0.034	--	--	--	--
Mode choice model																							
ASC for transit	-0.56	-0.53	5.4	0.066	0.069	-0.61	8.9	0.056	0.067	0.87	-0.22	60.0	0.053	0.057	3.59	-0.28	50.5	0.084	0.066	-0.24	56.9	0.053	0.056
ASC for walk	1.56	1.47	5.9	0.120	0.114	1.17	24.8	0.093	0.096	2.01	1.02	34.6	0.097	0.095	3.20	1.37	12.1	0.230	0.114	0.96	38.2	0.090	0.092
Mean of travel time coefficient γ_{TT^*}	-1.00	-0.91	8.9	0.048	0.045	-0.76	23.5	0.032	0.028	2.95	-0.67	32.7	0.029	0.028	4.34	-0.91	8.6	0.107	0.046	-0.61	38.9	0.020	0.019
SD of travel time coefficient γ_{TT^*}	0.19	0.19	0.5	0.020	0.016	--	--	--	--	--	0.09	50.2	0.016	0.014	3.92	0.17	8.8	0.053	0.017	--	--	--	--
Cost coefficient (β_C)	-0.25	-0.23	6.1	0.013	0.014	-0.21	15.5	0.011	0.011	1.22	-0.19	25.6	0.009	0.009	4.01	-0.25	0.9	0.025	0.014	-0.17	31.3	0.009	0.008
Mean of APB, FSSE, ASE	--	--	7.0	0.053	0.051	--	18.2	0.048	0.050	--	--	40.6	0.041	0.041	--	--	16.2	0.099	0.051	--	41.3	0.043	0.044

Between the *ICSV-RC* and *ML-SV* models, both of which accommodate both the sources of variability in consideration, the latter provides a slightly better fit. Since the two-step estimation using maximum likelihood techniques involves separate optimization of likelihoods in two different steps (the integrated estimation optimizes the joint likelihood), the two-step estimation is likely to show better fit than that from the simultaneous estimation approach, unless there is a large enough bias in parameter estimates due to endogeneity issues in the two-step approach (see Vij and Walker, 2016 for a similar finding and a discussion in the context of *ICLV* models). If there is no strong reason for endogeneity between the two steps conditional on the same travel time distribution entering the measurement equations in both the steps, it might be easier for the analyst to enhance the characterization of the travel time distribution using more advanced approaches while using sequential estimation (*i.e.,* the *ML-SV* approach) to estimate the relevant parameters. Of course, the standard errors may have to be corrected to address loss in efficiency.

Finally, during the estimation of the *ICSV-RC* model on each of the 200 datasets, we explored different sets of starting values for the parameters. For each dataset, the *ICSV-RC* model converged to the same maximum likelihood parameter estimates regardless of the starting parameter values employed in estimation. This pattern indicates that the *ICSV-RC* model did not encounter a flat likelihood surface at the maximum likelihood values of the parameters. That is, the *ICSV-RC* model can be used to simultaneously identify stochasticity in alternative attributes and their coefficients – if data are available on attribute measurements and traveller choices.

4.2 Simulation experiment Set II: Mode choice setting with lognormal distribution assumption for bus travel time and lognormal coefficient on travel time

For this set of experiments, we modified the simulations conducted for the mode choice setting discussed in Set I by assuming a lognormal distributed bus travel time and a lognormal distributed coefficient on travel time. To simulate bus travel times, the same equation ${TT_b}^* = \theta_b d_b$ was used, and θ_b was assumed to follow lognormal distribution whose underlying normal distribution location parameter is 0.49 min/km and scale parameter is 0.25. The assumptions for generating other exogenous variables remained unchanged. Next, the travel time coefficient (γ_{TT^*}) was assumed to follow a lognormal distribution with the underlying normal location parameter -1.00 and scale parameter 0.05 (specifically, the negative of the values drawn from this distribution were used for the travel time coefficient). True values assumed for other parameters were the same as those in Section 4.1.1. Using these assumptions, 200 mode choice datasets, each comprising 5,000 trips were simulated. For each of these trips from the 200 datasets, a single measurement of bus travel time was simulated by adding a normal distributed measurement error to the simulated value of ${TT_b}^*$. The standard deviation of this measurement error was assumed to be 0.95, as earlier.

The simulation results for this set of experiments (Table 2) indicate accurate and precise recovery of parameters for both the travel time and the mode choice components in the *ICSV-RC* model. Further, the APB values of the *ICSV-RC* model are lower than those of the *ICSV* and the *ML-RC* models. These findings are similar to those from the experiments in Set I. Next, note that the mean (across 200 datasets) estimate of the location parameter for the coefficient on travel time is biased away from zero in the *ML-RC* model when compared to that in the *ICSV-RC* model. On the other hand, the mean estimate of the scale parameter for this coefficient is biased toward zero. This contrasts with the finding in the context of experiments in Set I, where both location and scale parameters estimated from the *ML-RC* model were biased toward zero. This finding is in line with our theoretical discussion in Section 3 on the repercussions of ignoring stochasticity in variables with log-normal distribution – only one or both of the location and scale parameters might be biased toward zero, or one of the two parameter estimates gets biased toward zero while the other gets biased away from zero. All other parameter estimates in the *ML-RC* mode choice utility functions demonstrate a bias toward zero when compared to those in the *ICSV-RC* model. In addition, the magnitude of bias in the *ICSV* model is lower than that in the *ML-RC* model.

The simulation results for the *ICSV-RC* and *ML-RC* models in Table 1 and Table 2 highlight another finding. Specifically, note from Table 1 that the trace of the covariance matrix of coefficients for the choice model component was 0.258 when stochasticity in bus travel time was incorporated in the model (i.e., the *ICSV-RC* model) and 0.203 when stochasticity in bus travel time was present but ignored in the model (i.e., the *ML-RC* model). The corresponding trace values for the *ICSV-RC* and *ML-RC* models in Table 2 are 0.271 and 0.220, respectively. These results indicate that the standard errors are underestimated using the *ML-RC* model that ignores stochasticity in an exogenous variable. This is because the standard errors of coefficients in the *ML-RC* model are predicated on the assumption that the observations on the stochastic exogenous variable will remain the same in repeated samples. Given a stochastic exogenous variable, this will not be the case, implying that the standard errors will, in general, be underestimated using a framework such as the *ML-RC* (leading to potentially incorrect inferences).

Table 2 Simulation evaluation results for the mode choice setting (Set II): Lognormal distribution assumption for travel time and its coefficient

	True value	***ICSV-RC*** **model parameter estimates**				***ICSV*** **model parameter estimates**					***ML-RC*** **model parameter estimates**				
		Mean	APB	FSSE	ASE	Mean	APB	FSSE	ASE	t-stat. for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ICSV}$	Mean	APB	FSSE	ASE	t-stat. for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ML-RC}$
Bus travel time model															
Inverse speed θ_b: Location parameter	0.49	0.44	10.48	0.008	0.004	0.44	11.11	0.008	0.003	0.49	--	--	--	--	--
Inverse speed θ_b: Scale parameter	0.25	0.22	10.07	0.013	0.002	0.22	11.71	0.009	0.002	1.58	--	--	--	--	--
SD of measurement error	0.95	0.95	0.43	0.006	0.006	0.95	0.42	0.007	0.006	0.02	--	--	--	--	--
Mean APB, FSSE, ASE	--	--	6.99	0.009	0.004	--	7.74	0.008	0.004	--	--	--	--	--	--
Mode choice model															
ASC for transit	-0.56	-0.56	0.07	0.067	0.070	-0.68	22.25	0.081	0.064	1.32	-0.32	42.69	0.046	0.053	2.74
ASC for walk	1.56	1.41	9.35	0.176	0.122	1.22	21.61	0.141	0.112	1.02	0.64	59.07	0.097	0.101	4.38
Travel time coefficient γ_{TT^*} - Location parameter	-1.00	-0.98	1.72	0.051	0.041	-0.97	2.53	0.069	0.037	0.12	-1.39	38.94	0.028	0.033	9.46
Travel time coefficient γ_{TT^*} - Scale parameter	0.05	0.05	1.90	0.029	0.027	0.00	--	--	--	--	0.00039	99.21	0.0001	0.026	1.31
Cost coefficient (β_C)	-0.25	-0.26	5.34	0.013	0.011	-0.27	6.78	0.018	0.011	0.19	-0.18	28.01	0.005	0.007	7.57
Mean of APB, FSSE, ASE	--	--	6.68	0.067	0.054	--	13.29	0.077	0.056	--	--	53.58	0.035	0.044	--

5 EMPIRICAL ANALYSIS

In this section, we present an empirical analysis for a joint analysis of route-level travel time and route choice while considering both stochasticity in network travel times and random heterogeneity in sensitivity to travel time. This empirical analysis is focused more on corroborating the findings from the earlier sections than on the substantive aspect of route choice analysis itself.

5.1 Empirical data

The main source of empirical data for this analysis, provided by the American Transportation Research Institute (ATRI) is a large truck-GPS dataset of about 96 million GPS traces in the state of Florida, USA (Pinjari *et al*., 2015). The raw data were first converted into a database of truck trips by Thakur *et al*. (2015) using GPS-to-trip conversion algorithms. For these trips, the travelled routes were not readily observable in the form of network links and nodes traversed between the OD locations. The raw GPS data was map-matched to the roadway network to derive the travelled routes using a high-resolution roadway network obtained from the Florida Department of Transportation (FDOT) (Tahlyan *et al*., 2017). Such truck route choice data were generated for a total of 8211 truck trips in the state of Florida.

For all the 8211 trucks trips used in this study, route choice sets were generated by Tahlyan and Pinjari (2020) using the Breadth First Search-Link Elimination (BFS-LE) algorithm proposed by Rieser-Schüssler *et al.* (2013). [9] For each trip, the BFS-LE algorithm was run to generate up to 16 unique route choice alternatives. For some of these trips, the chosen route was included as an additional choice alternative since the choice set did not include the chosen route completely. Next, for all route alternatives of each of the 8211 truck trips, route attributes such as the total route length, lengths on different types of roads, number of intersections, and the proportion of toll road length were derived. In addition, to account for the degree of overlap of a route with other routes in the choice set for that same OD pair, a path-size attribute (Ben-Akiva and Bierlaire, 1999) was

[9] The BFS-LE is a deterministic link elimination approach based on a repeated least cost path search, where links on the current shortest path are eliminated, one by one, to find subsequent least cost paths. Hence, it is well-suited for extracting routes from large-scale, high-resolution networks. The primary difference between this algorithm when compared against other link-elimination approaches is that it uses a tree structure in which each node is a network. Starting initially with the original network (which is the root node of the tree), any unique network obtained after the elimination of a link from a current least cost path is a node of the tree, given that the network offers at least one feasible route for the OD pair under consideration.

computed, where a greater path-size value indicates a smaller extent of overlap and a path-size value of one indicates no overlap.[10]

The same GPS data were used to extract measurements of travel times for each of the 8211 chosen routes. Among the chosen routes, the number of available travel time measurements varied from one to as many as ten or more, although more than 70% of the chosen routes had only one or two measurements. Non-chosen route alternatives did not have travel time measurements.

It is worth noting here that travel time measurements were available for only chosen routes. However, the chosen routes across all the trips in the dataset provided good spatial coverage of the network and the different types of links in the roadway hierarchy (the coverage was assessed based on a visual examination of all the chosen routes overlaid on the complete network). Further, the chosen routes overlapped to some extent with the non-chosen routes, which is a common phenomenon in route choice sets. Besides, the data were sampled from different days across four different months (October 2015, December 2015, April 2016, June 2016) and different times of the day. Therefore, a bias due to using only the chosen routes' travel time data is less likely in the current empirical study. However, if the chosen routes in the estimation dataset lacked adequate spatiotemporal coverage or if systematically more or less congested parts of the network were not represented in the observed route choice data, the model parameter estimates would likely be biased. This is because the travel time model parameter estimates would be based on data that might not provide adequate spatiotemporal coverage of travel conditions. To address this issue, one needs to explore other sources of travel time measurements for travel conditions that are not well-represented in the observed route choice data. Since emerging travel time data sources such as GPS probe data are typically very large and provide a very good coverage of the network, the use of travel time measurements of only chosen routes is not likely to be a major concern in other such applications.

5.2 Empirical results and findings

Table 3 presents the parameter estimates for the empirical route choice models estimated in this study – *ICSV-RC*, *ICSV*, *ML-RC*, *ML-SV* and *ML-EC*. In all these models, we included error

---

[10] The path-size variable for a route i is defined as: $PS_i = \sum_{a \epsilon \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \epsilon C_n} \delta_{aj}}$, where Γ_i is the set of all links in path/route i between the OD pair n, l_a is the length of link a, L_i is the length of path i, C_n is the choice set of route alternatives between the OD pair n. δ_{aj} is equal to 1 if a route $j \epsilon C_n$ uses link a and 0 otherwise.

components to consider correlations among route-specific utility functions due to unobserved factors. Due to the importance of such error correlations in route choice settings, we do not report a simple *MNL* model. Next, we briefly discuss the empirical results from the *ICSV-RC* model and compare them with those of other models to evaluate the importance of accommodating both sources of stochasticity discussed earlier.

5.2.1 Empirical results from the ICSV-RC model

The columns in Table 2 under the title "*ICSV-RC* model" present the parameter estimates for the *ICSV-RC* model where both the travel time and its coefficient are specified as random. In this model, the random coefficients in the stochastic travel time function (Equation (3)) were specified as normally distributed. Other distributional assumptions could be made in this regard, such as a truncated normal or a shifted lognormal; however, we used the normal distribution specification as an initial effort to disentangle variability in travel time from that in its coefficient (which is also assumed to be normally distributed).

As can be observed from the parameter estimates of the stochastic travel time equation, the estimates for mean inverse speeds (in minutes per mile) and the corresponding standard deviations are in increasing order from interstates to local roads. This is intuitive given that interstates figure at the top in the hierarchy of functional classification of roadways. Further, the probability of zero or negative values of the mean inverse speeds is zero for all practical purposes (less than 8.6×10^{-3} for minor arterials and of the order of 10^{-6} or lesser for other roadway functional classes). The value of mean junction-crossing time at turns (0.194 minutes per turn) also turned out to be statistically significant.

As discussed in Section 2, the measurement equation for the travel time model includes a measurement error term. The standard deviation estimate in the measurement equation for travel time suggests a significant error in the measurement or extraction of travel time using GPS data.

Table 3 Empirical results for the route choice setting

	ICSV-RC model		ICSV model			ML-RC model			ML-SV model		ML-EC model	
	Par. est.	Std. error	Par. est.	Std. error	t-stat for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ICSV}$	Par. est.	Std. error	t-stat for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ML-RC}$	Par. est.	Std. error	Par. est.	Std. error
Structural eqn. for stochastic travel time												
Interstate highway length - mean parameter	0.955	0.0017	0.945	0.0017	--	--	--	--	0.965	0.0019	--	--
Major arterial length - mean parameter	1.284	0.0050	1.322	0.0056	--	--	--	--	1.277	0.0069	--	--
Minor arterial length - mean parameter	1.599	0.0120	1.709	0.0156	--	--	--	--	1.425	0.0192	--	--
Collector street length - mean parameter	1.924	0.0199	2.180	0.0267	--	--	--	--	1.690	0.0318	--	--
Local road length - mean parameter	2.784	0.0398	2.851	0.0458	--	--	--	--	2.881	0.0654	--	--
Total number of junctions - mean parameter	0.194	0.0148	0.067	0.0155	--	--	--	--	0.275	0.0239	--	--
Interstate highway length - SD parameter	0.069	0.0006	0.060	0.0006	--	--	--	--	0.068	0.0069	--	--
Major arterial length - SD parameter	0.212	0.0042	0.235	0.0041	--	--	--	--	0.174	0.0055	--	--
Minor arterial length - SD parameter	0.510	0.0038	0.573	0.0050	--	--	--	--	0.522	0.0062	--	--
Collector street length - SD parameter	0.407	0.0153	0.559	0.0176	--	--	--	--	0.498	0.0280	--	--
Measurement eqn. for travel time												
SD of measurement error in GPS data	3.603	0.0024	3.597	0.0025	--	--	--	--	3.597	0.0024	--	--
Route choice utility functions												
Mean of route-level travel time coefficient	-1.243	0.0470	-0.475	0.0090	16.05	-1.072	0.0346	2.94	-1.276	0.0535	-0.449	0.0064
SD of route-level travel time coefficient	0.871	0.0338	0.000	--	--	0.678	0.0253	4.56	0.979	0.0402	--	--
Natural logarithm of path size	-2.340	0.0880	-2.048	0.0597	2.75	-1.119	0.6331	1.91	-3.058	0.0927	-1.014	0.5114
Proportion of tolled portion on the route	-7.644	1.0444	-7.214	0.7804	0.33	-5.245	0.0640	2.29	-7.029	0.9980	-3.637	0.0892
Error components in utility functions												
SD of error component on square root of route length on Interstate 75 in Florida	5.558	0.2522	2.945	0.1196	9.35	4.257	0.2172	3.91	6.285	0.2890	2.994	0.0923
SD of error component on square root of route length on Polk Parkway in Florida	3.403	0.2813	2.724	0.2548	1.79	3.127	0.3568	0.61	4.135	0.4789	2.386	0.2092

Moving on to the route-choice model component, it is notable that in addition to allowing the estimation of random coefficients in the stochastic travel time equation (i.e., stochasticity in travel time), the model allows the estimation of a random coefficient on travel time in its route choice utility component. Specifically, the mean and standard deviation parameter estimates of γ_{TT^*} (see the coefficient on route-level travel time distribution and its standard deviation in Table 2) are statistically significant and reasonable, with more than 92% of the population having a negative value for γ_{TT^*}.

In the remainder of the route choice utility function, the coefficient for the natural logarithm of path size has a negative sign, which is expected because routes with higher overlap would each have a lower probability of being chosen than the probability of all of them being chosen. Further, the coefficient on the proportion of tolled roads on a route is negative, indicating lower utilities for routes having greater proportions of tolled lengths, *ceteris paribus*. In addition, the error components specified in the utility functions to capture inter-route correlations are statistically significant.

5.2.2 Empirical results from the ICSV model

Now, we turn to the set of parameter estimates for the *ICSV* model in Table 2, where the travel time coefficient was estimated as a fixed parameter. The parameter estimates for the choice model component in this model are lower in magnitude than those in the *ICSV-RC* model. These differences in the parameter estimates between the two models, as evident from the corresponding t-statistic values reported in the table under the column '*t-stat for* $H_0: \hat{\beta}_{ICSV-RC} = \hat{\beta}_{ICSV}$', are statistically significant. As discussed in Section 3, ignoring randomness in the coefficient on the stochastic travel time has likely led to a systematic bias in the parameter estimates of the choice model.

5.2.3 Empirical results from mixed logit models

Let us now examine the results for the *ML-RC* model, which ignores the stochasticity in travel time. Similar to the *ICSV* model, the *ML-RC* route choice model parameter estimates for the mean and standard deviation coefficients on route-level travel time are lower in magnitude than those in the *ICSV-RC* model. This finding, once again, corroborates our claims from Section 3. In particular, the statistically significant underestimation of the two primary model parameters under scrutiny (the mean and standard deviation of the coefficient on travel time) when travel time stochasticity is ignored highlights the drawbacks of using conventional mixed logit models in settings that involve random variables such as travel time as well as randomness in the sensitivity to such variables.

The *ML-EC* model's coefficient on route-level travel time also shows a bias toward zero. Since this model ignores both sources of variability, the bias in its parameter estimates is greater than that in the *ICSV* model or the *ML-RC* model.

Next, recall that the *ML-SV* model differs from the *ICSV-RC* model in that the distribution parameters of travel time are estimated a priori from a predecessor travel time model. As such, the estimates presented for the bus travel time model (under the column '*ML-SV model*' in

Table 3) are based on a separate earlier estimation. Comparing the route choice components of the *ICSV-RC* model and those of the *ML-SV* model, one can observe that the parameter estimates are quite close between the two models. This is because both these models consider stochasticity in travel times and a random coefficient on travel time. However, as discussed earlier, the standard errors of the parameter estimates in the *ML-SV* model are higher than those in the *ICSV-RC* model.

5.2.4 Goodness-of-fit in estimation and validation samples

To assess the goodness-of-fit of the various empirical models estimated in this study, we conducted a five-fold validation. That is, from the full dataset of 8,211 trips available for the empirical study, we randomly drew five estimation samples of 6,453 trips and estimated all the above-discussed models. For each of the five estimation samples of 6,453 trips, the remaining 1,758 trips were kept aside for validation purposes. Subsequently, we computed the goodness-of-fit metrics shown in Table 4 for all five sets of estimation and validation samples (for each set of estimation and validation samples, the parameter estimates from the corresponding estimation sample were used). These metrics include log-likelihood at convergence for the integrated models, log-likelihood at convergence for only the route choice model component, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for the route choice model component, and adjusted McFadden's Rho-square for the route choice model component. Average values of these metrics (averaged across the five sample datasets) are reported in Table 3 – separately for the estimation samples (in the upper set of rows) and the validation samples (in the lower set of rows).

Table 4 Goodness-of-fit metrics in estimation and validation samples

Goodness-of-fit measures in estimation samples	***ICSV-RC model***	***ICSV model***	***ML-RC model***	***ML-SV model***	***ML-EC model***
Log-likelihood at convergence for both travel times and route choices	-215846.07	-216424.19	--	-209100.62	--
Log-likelihood of only route choice component	-6464.24	-6935.97	-6554.99	-6458.20	-7178.41
No. of parameters in the full model	16	15	6	16	5
No. of parameters estimated in the choice model	16	15	6	16	5
AIC for the full model	431724.14	432878.38	13121.98	418233.24	14366.82
AIC for choice model component	12960.47	13901.94	13121.98	12948.40	14366.82
BIC for the full model	431832.50	432979.96	13162.61	418341.60	14400.68
BIC for choice model component	13068.83	14003.52	13162.61	13056.75	14400.68
Adjusted Rho-squared for the full model	0.756	0.755	0.589	0.763	0.550
Adjusted Rho-squared for the choice model	0.594	0.564	0.589	0.594	0.550
Goodness-of-fit measures in validation samples using parameters from estimation samples	***ICSV-RC model***	***ICSV model***	***ML-RC model***	***ML-SV model***	***ML-EC model***
Log-likelihood at estimated parameter values for both travel times and route choices	-62750.09	-62889.24	--	-60589.32	--
Log-likelihood of only route choice component	-1687.70	-1840.94	-1701.32	-1683.49	-1896.91
AIC for the full model	125532.20	125808.48	3414.65	121210.64	3803.82
AIC for choice model component	3407.39	3711.89	3414.65	3378.99	3803.82
BIC for the full model	125619.74	125890.56	3447.48	121298.19	3831.18
BIC for choice model component	3494.94	3793.97	3447.48	3486.54	3831.18
Adjusted Rho-squared for the full model	0.753	0.752	0.622	0.761	0.578
Adjusted Rho-squared for the choice model	0.622	0.589	0.622	0.623	0.578

As can be observed from the table, the choice model components of both *ICSV-RC* and the two-step *ML-SV* models provide better goodness of fit measures than those of *ICSV*, the *ML-RC*, and the *ML-EC* models. The same trend can be observed in both estimation and validation samples. It is an expected result that the models that incorporate both stochasticity in travel times and randomness in the coefficient on travel time provide a better fit than other models that ignore one or both sources of variability. However, it is interesting to note that the loss in fit is greater when the variability in the random coefficient on travel time is ignored (*ICSV* model) than when stochasticity in travel time is ignored (*ML-RC* model).

Between the *ICSV-RC* and *ML-SV* models, both of which accommodate both the sources of variability in consideration, the latter provides a slightly better fit. A plausible reason for this has been discussed in Section 4.2 in the context of simulation results. This result, combined with the similarity of the parameter estimates between the two models, opens scope for enhancing the characterization of the travel time distribution using more advanced approaches while using sequential estimation to estimate the relevant parameters.

6 SUMMARY AND CONCLUSIONS

In this study, we formulate a choice modelling framework that allows the analyst to accommodate stochasticity in explanatory variables and random coefficients on such variables. Specifically, we develop an integrated choice and stochastic variable modelling framework with random coefficients (i.e., an *ICSV-RC* framework) to disentangle travel time variability from unobserved heterogeneity in response to travel time in travel choice models. The *ICSV-RC* model allows the identification of both the sources of variability – stochastic explanatory variables (such as travel time) and random coefficients on those variables – due to its ability to bring together travel choice data and measurement data for the stochastic variables. The measurement data of the stochastic variables (travel time) allows the estimation of parameters for the stochastic variables and the travel choice data allows the estimation of random coefficients on the stochastic variables. In addition, we show that ignoring either source of stochasticity – stochasticity in alternative attributes or heterogeneity in response to the attributes – results in models with inferior goodness-of-fit and a systematic bias in all parameter estimates. If the stochasticity in an alternative attribute is ignored and the random coefficient on that attribute is distributed such that the location and scale parameter are additively separable (e.g., normal distributed random coefficient), we show that the estimates of both location and scale parameters of the random coefficient would be biased toward zero. Furthermore, ignoring stochasticity in an alternative attribute, when stochasticity is present, leads to underestimation of the standard errors. We demonstrate such repercussions of ignoring stochastic explanatory variables using simulation experiments in two distinct choice settings – one involving labelled mode choice alternatives and the other involving unlabelled route choice alternatives. Furthermore, we applied the *ICSV-RC* model to an empirical analysis of truck route choice in Florida, USA. The integrated model was found to successfully disentangle stochasticity in route-level travel time from heterogeneity in response to travel time. Simpler versions of the model that ignore either stochasticity in travel time or impose a deterministic coefficient on travel time had inferior goodness-of-fit and showed a bias in the parameter estimates. These results highlight the importance of accounting for both sources of variability.

The *ICSV-RC* methodology in this paper overcomes the limitation of mixed logit/probit models used to accommodate random coefficients on deterministic explanatory variables and the limitation of *ICLV* models used to accommodate latent (stochastic) variables with deterministic coefficients. It has hitherto been believed that identification of both these sources of variability – stochastic attributes and random coefficients on those attributes – is very difficult, if not impossible (Diáz *et al*., 2015). In addition to the *ICSV-RC* model that involves a simultaneous estimation, we discuss a sequential, two-step estimator (*ML-SV* model) that optimizes the likelihoods of observed measurements of stochastic variables and that of observed travel choices separately. The sequential estimator can be useful in situations where the specifications for the stochastic attributes' distribution is complex enough to make it difficult for simultaneous estimation. Going forward, we hope that such models – which use multiple sources of data – will help increase the recognition of stochastic variables and random coefficients in choice models.

Some limitations of this study offer scope for further research. First, in the empirical analysis we used the normal distributional assumption for travel time as a first step to address the core idea that the paper puts forth – the identification of the two sources of variability. The authors recognize that a normal (or truncated normal) distribution for travel time would imply, for a given distance, a reciprocal normal (reciprocal of truncated normal) distribution for travel speed, which leads to theoretically undefined mean and variance parameters (unless the normal distribution is bounded strictly away from zero). Therefore, it is important to explore alternative distributional forms for the stochastic travel time variable. Second, we used the maximum simulated likelihood estimation method in this study. The estimation time for each dataset was about two hours on a workstation-grade computer. The exploration of alternative estimation methods for the *ICSV-RC* model is a fruitful research avenue. Finally, the current study focuses on the stochasticity of alternative attributes, which vary across choice alternatives. It will be useful to explore avenues to identify the stochasticity of choice environment variables that do not vary across choice alternatives. A recent study by Nirmale and Pinjari (2023) is a step forward in this direction.

ACKNOWLEDGMENTS

This research was undertaken as part of a SPARC project #256 funded by the Indian Ministry of Education. In addition, the second author acknowledges support from the Satish Dhawan-loE-Visiting Chair Professor position at the Indian Institute of Science. The authors are thankful to the American Transportation Research Institute (ATRI) for providing the truck-GPS data and to Divyakant Tahlyan for providing the processed data for the empirical analysis of route choice presented in this study. Two anonymous reviewers provided helpful comments on an earlier version of this paper.

Appendix A: Simulation evaluation for the route choice setting (Set III)

In addition to the mode choice simulation experiments presented in Section 4.1 of the paper, we conducted a second set of simulation experiments for a route choice setting with unlabelled alternatives. To keep the synthetic data realistic (akin to that from a real road network), we drew data on exogenous variables and choice sets from the empirical route choice data used in Section 5 of the study. The true values of the parameters assumed to simulate the data are taken from the parameter estimates reported for the empirical *ICSV-RC* model in Table 2. Using these parameters and the exogenous variable data, we generated 200 route choice datasets, each with a sample size of 2000 trips for the *ICSV-RC* model. For each of the 2000 trips in the 200 datasets, we generated 10 travel time measurements for the simulated chosen route. We assumed that the non-chosen routes would not have travel time measurements.

We estimated all the models discussed in Section 2 on all the 200 simulated datasets and then computed the parameter recovery metrics discussed in Section 4 (i.e., APB, ASE, and FSSE). These metrics are reported in Table A1 below. As can be observed from this table, the trends in overall parameter recovery and bias in parameter estimates are similar to those observed in Section 4.1 for the mode choice context. These findings thus underscore the need for a model framework such as the *ICSV-RC* model to accommodate both stochastic variables and random coefficients when compared to the other models typically used in the literature.

Table A1 Simulation evaluation results for the route choice setting: Normal distribution assumption for stochastic route-level travel time and its coefficient

Variable description	True value	***ICSV-RC* model parameter estimates**				***ICSV* model parameter estimates**					***ML-RC* model parameter estimates**				
		Mean	APB	FSSE	ASE	Mean	APB	FSSE	ASE	t-stat. for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ICSV}$	Mean	APB	FSSE	ASE	t-stat. for H_0: $\hat{\beta}_{ICSV-RC} = \hat{\beta}_{ML-RC}$
Structural eqn. for stochastic travel time															
Interstate highway length - mean parameter	0.955	0.955	0.06	0.0034	0.0026	0.953	0.16	0.0038	0.0026	--	--	--	--	--	--
Major arterial length - mean parameter	1.284	1.284	0.02	0.0172	0.0149	1.283	0.07	0.0164	0.0071	--	--	--	--	--	--
Minor arterial length - mean parameter	1.599	1.596	0.17	0.0337	0.0238	1.597	0.11	0.0331	0.0139	--	--	--	--	--	--
Collector street length - mean parameter	1.924	1.919	0.25	0.0475	0.0402	1.907	0.89	0.0513	0.0201	--	--	--	--	--	--
Local road length - mean parameter	2.784	2.802	0.63	0.1065	0.0919	2.816	1.13	0.1085	0.0909	--	--	--	--	--	--
Total number of turns - mean parameter	0.194	0.198	2.20	0.0320	0.0271	0.203	4.95	0.0301	0.0270	--	--	--	--	--	--
Interstate highway length - SD parameter	0.069	0.064	7.85	0.0162	0.0170	0.065	5.86	0.0036	0.0017	--	--	--	--	--	--
Major arterial length - SD parameter	0.212	0.214	0.99	0.0148	0.0110	0.228	3.20	0.0151	0.0045	--	--	--	--	--	--
Minor arterial length - SD parameter	0.510	0.521	2.22	0.0293	0.0208	0.513	0.55	0.0251	0.0075	--	--	--	--	--	--
Collector street length - SD parameter	0.407	0.398	2.13	0.0369	0.0324	0.391	3.96	0.0309	0.0117	--	--	--	--	--	--
Measurement eqn. for travel time															
SD of measurement error in GPS data	3.603	3.790	5.19	0.0256	0.0266	3.783	4.99	0.0018	0.0240	--	--	--	--	--	--
Mean of APB, FSSE, ASE values	--	--	1.97	0.0330	0.0280	--	2.09	0.0318	0.0187	--	--	--	--	--	--
Model (utility functions) for route choice															
Coefficient on travel time - mean parameter	-1.243	-1.088	12.50	0.0652	0.0625	-0.875	29.61	0.0026	0.0063	3.40	-0.977	21.45	0.0515	0.0506	1.38
Coefficient on travel time - SD parameter	0.871	0.790	9.31	0.0528	0.0490	--	--	--	--	--	0.658	24.50	0.0400	0.0407	2.09
Natural logarithm of path size	-2.340	-1.984	15.23	0.1523	0.1478	-1.569	32.96	0.1080	0.1500	1.98	-0.894	61.82	0.1520	0.1013	6.08
Proportion of tolled portion on the route	-7.644	-6.222	18.60	1.3051	1.3013	-4.599	39.84	1.0560	0.9860	0.99	-4.496	41.19	0.9553	0.9319	1.07
Error components for inter-route correlations															
SD of error component on square root of route length on interstate 75 in Florida	5.558	4.718	15.10	0.5196	0.3796	2.869	48.38	0.3300	0.5680	2.74	4.372	21.33	0.4884	0.373	0.71
SD of error component on square root of route length on interstate 75 in Florida	3.403	2.972	12.66	0.6399	0.5124	1.587	53.37	0.6690	0.8190	1.43	2.447	28.09	0.5262	0.4806	0.78
Mean of APB, FSSE, ASE values	--	--	13.90	0.4560	0.4090	--	40.83	0.4331	0.5059	--	--	33.06	0.3689	0.3296	--

Table A1 (Contd.) Simulation evaluation results for the route choice setting: Normal distribution assumption for stochastic route-level travel time and its coefficient

Variable description	True value	***ML-SV model* parameter estimates**				***ML-EC* model parameter estimates**			
		Mean	APB	FSSE	ASE	Mean	APB	FSSE	ASE
Structural eqn. for stochastic travel time									
Interstate highway length - mean parameter	0.955	0.954	0.12	0.0036	0.0030	--	--	--	--
Major arterial length - mean parameter	1.284	1.272	0.93	0.0146	0.0108	--	--	--	--
Minor arterial length - mean parameter	1.599	1.505	5.82	0.0289	0.0229	--	--	--	--
Collector street length - mean parameter	1.924	1.839	4.39	0.0500	0.0386	--	--	--	--
Local road length - mean parameter	2.784	2.861	2.74	0.1137	0.1376	--	--	--	--
Total number of turns - mean parameter	0.194	0.256	31.87	0.0465	0.0433	--	--	--	--
Interstate highway length - SD parameter	0.069	0.065	6.53	0.0033	0.0019	--	--	--	--
Major arterial length - SD parameter	0.212	0.214	0.95	0.0152	0.0062	--	--	--	--
Minor arterial length - SD parameter	0.510	0.497	2.63	0.0198	0.0122	--	--	--	--
Collector street length - SD parameter	0.407	0.393	3.53	0.0334	0.0268	--	--	--	--
Measurement eqn. for travel time						--	--	--	--
SD of measurement error in GPS data	3.603	3.780	4.90	0.0016	0.0240	--	--	--	--
Mean of APB, FSSE, ASE values	--	--	5.85	0.0301	0.0298	--	--	--	--
Model (utility functions) for route choice									
Coefficient on travel time - mean parameter	-1.243	-1.114	10.43	0.0613	0.0681	-0.446	64.11	0.0089	0.0102
Coefficient on travel time - SD parameter	0.871	0.766	12.05	0.0497	0.0548	--	--	--	--
Natural logarithm of path size	-2.340	-1.933	17.43	0.1429	0.1522	-1.727	26.21	0.4893	0.6653
Proportion of tolled portion on the route	-7.644	-5.897	22.85	1.4252	1.2845	-4.230	44.66	0.0269	0.5660
Error components for inter-route correlations									
SD of error component on square root of route length on interstate 75 in Florida	5.558	5.558	15.07	0.4383	0.4495	2.107	62.10	0.1025	0.1996
SD of error component on square root of route length on interstate 75 in Florida	3.403	3.403	12.60	0.5066	0.6053	2.569	24.48	0.3669	0.2056
Mean of APB, FSSE, ASE values	--	--	15.07	0.4373	0.4357	--	44.31	0.1989	0.3293

Appendix B: Simulation evaluation of the effects of incorrect distributional assumptions (Set IV)

To complement the simulation experiments conducted in Section 4 and Appendix A, we conducted additional experiments to assess the impact of incorrect assumptions for the distributions of the stochastic variable and its random coefficient. Specifically, we investigated the following two cases: (1) the true data generation process (DGP) involved lognormal distributed bus travel times and lognormal coefficient on travel times while the estimation was carried out assuming normal distributions for both bus travel times and travel time coefficient, and (2) the true DGP involved travel times that are free of measurement errors (due to absence of stochasticity) and a lognormal coefficient on travel time whereas the estimation was carried out assuming normal distributions for both bus travel times and travel time coefficient.

In the simulation design for the first case, a lognormal distribution was assumed for the inverse speed of bus mode with the underlying normal location parameter 0.49 min/km and scale parameter 0.25. The travel time coefficient was assumed to follow a lognormal distribution with the underlying normal location parameter -0.36 and scale parameter 0.22 (specifically, the negative of the values drawn from this distribution were used for the travel time coefficient).

For the second case, the inverse speed of bus was assumed to be 1.5 min/km. The travel time coefficient was assumed to follow a lognormal distribution with the underlying normal location parameter -1.00 and scale parameter 0.05 (the negative of the values drawn from this distribution were used for the travel time coefficient). Assumptions for generating other exogenous variables and the true values assumed for all other parameter values remained the same as in Section 4.1.

For each of the two cases discussed above, 200 datasets were generated, each comprising 5000 trips. The simulation results for the *ICSV-RC* model for the first case are presented in Table B1. These results indicate large APB values for all the parameters. However, since the distributions in the true DGP are different from the ones assumed in model estimation, comparing the true parameters vis-à-vis the recovered parameters in terms of bias is not helpful. Therefore, we compared the values-of-time (VoT) between the true and estimated models.

Table B1 Simulation evaluation results for the mode choice setting with lognormal-lognormal assumption for travel time and its random coefficient in the DGP and normal-normal assumption in estimation

	True value	***ICSV-RC* model results**			
		Mean	APB	FSSE	ASE
Bus travel time model					
Inverse speed θ_b: Location parameter	0.49	1.56	221.73	0.011	0.004
Inverse speed θ_b: Scale parameter	0.15	0.20	31.08	0.007	0.003
SD of measurement error	0.95	0.95	0.46	0.005	0.008
Mean APB, FSSE, ASE	--	--	84.42	0.007	0.005
Mode choice model					
ASC for transit	-0.56	-0.61	8.42	0.121	0.067
ASC for walk	1.56	1.23	21.03	0.136	0.102
Travel time coefficient γ_{TT^*} - Location parameter	-0.36	-0.49	35.57	0.037	0.018
Travel time coefficient γ_{TT^*} -Scale parameter	0.22	0.04	83.23	0.054	0.009
Cost coefficient (β_C)	-0.25	-0.18	26.98	0.012	0.009
Mean of APB, FSSE, ASE	--	--	35.05	0.072	0.041

From Table B1, the expected value of γ_{TT^*} in the true DGP = -0.71; standard deviation of γ_{TT^*} in the true DGP = 0.83; expected value of γ_{TT^*} in the estimated model = -0.61; standard deviation of γ_{TT^*} in the estimated model = 0.37. Thus, we obtain an expected VoT of 170.4 (INR/hour) in the true DGP (with lognormal distributions for bus travel time and random coefficient on travel time) and an expected VoT of 203.3 (INR/hour) for the estimated model (which incorrectly specifies normal distributions on both bus travel time and travel time coefficient). That is, in this case, the VoT is overestimated due to incorrect distributional assumptions.

The second case for this set of experiments involves measurement error-free travel times (due to absence of stochasticity in travel conditions) and lognormal coefficient on travel time in the DGP, whereas we assumed normal distributions during estimation. It can be observed from Table B2 that an incorrect specification of the distributions in the model estimation stage results in a high APB for the location and scale parameters of the coefficient on travel time. All other parameters are recovered well in terms of accuracy and precision. Notably, the variability in travel time (which is not present in the DGP) is estimated and its recovered value is close to 0 (the corresponding parameter estimate from most of the datasets was statistically insignificant). This indicates that our attempts to recover stochasticity in a variable when there was no stochasticity in the true DGP were innocuous and did not hurt the model. In this case too, examining only the APB of parameter estimates can be misleading because of different distribution assumptions in the true model and the estimated model. Hence, a comparison of the value-of-time metric is carried out.

Table B2 Simulation evaluation results for the mode choice setting with measurement error-free travel time and lognormal coefficient on travel time in the DGP and normal-normal assumption in estimation

	True value	***ICSV-RC* model results**			
		Mean	APB	FSSE	ASE
Bus travel time model					
Inverse speed θ_b: Location parameter	1.50	1.34	13.89	0.116	0.131
Inverse speed θ_b: Scale parameter	0.00	0.00052	--	0.003	0.003
SD of measurement error	0.95	0.95	0.07	0.006	0.006
Mean APB, FSSE, ASE	--	--	6.98	0.061	0.069
Mode choice model					
ASC for transit	-0.56	-0.53	5.89	0.069	0.061
ASC for walk	1.56	1.34	13.89	0.116	0.131
Travel time coefficient γ_{TT^*} - Location parameter	-1.00	-0.38	61.91	0.017	0.018
Travel time coefficient γ_{TT^*} -Scale parameter	0.05	0.02	54.63	0.022	0.022
Cost coefficient (β_C)	-0.25	-0.26	2.90	0.009	0.009
Mean of APB, FSSE, ASE	--	--	27.84	0.047	0.048

From the table above, the expected value of γ_{TT^*} in the true DGP = -0.37; standard deviation of γ_{TT^*} in the true DGP = 0.22, expected value of γ_{TT^*} in the estimated model = -0.68; standard deviation of γ_{TT^*} in the estimated model = 0.67. An expected VoT of 88.4 (INR/hour) is computed for the model in the true DGP while an expected VoT of 156.9 (INR/hour) is obtained for the estimated model. That is, the VoT is overestimated in the latter model due to incorrect distributional assumption on the travel time coefficient.

In summary, incorrect distributional assumptions in model estimation led to distorted values-of-time estimates, which can thereby lead to distorted policy analyses. It is worth noting here that the model is not harmed if the analyst estimates the stochasticity in an exogenous variable which is free of measurement errors.

Appendix C: Simulation evaluation of the effects of treating travel time as stochastic (and estimating its distribution) when there is no measurement error in travel time (Set V)

Here we discuss the effects of treating bus travel time as stochastic (and estimating the parameters of its distribution) when there is no measurement error (or stochasticity) in travel time in a mode choice setting. For this set of experiments, two specific cases were investigated: (1) bus travel time was free of measurement errors but its distribution was estimated (assuming that it is normal distributed), and the coefficient on travel time was assumed to be normal distributed, and (2) bus travel time was free of measurement errors but its distribution is estimated (assuming that it is lognormal distributed), and the coefficient on travel time was assumed to be lognormal distributed.

The simulation design for the first case assumed an inverse speed of bus as 1.5 min/km. The travel time coefficient was assumed to follow a normal distribution with mean -1.00 and SD 0.19. For the second case, the inverse speed of bus was assumed to be 1.63 min/km. The travel time coefficient was assumed to follow a lognormal distribution with the underlying normal distribution's location parameter -1.00 and scale parameter 0.05 (the negative of the values drawn from this distribution were used for travel time coefficient). The assumptions made for generating other exogenous variables and the parameter values remained the same as in Section 4.1.

For each of the above two cases discussed in this set of experiments, 200 datasets were generated, each comprising 5000 trips. The simulation results for each of these cases are presented in Table C1 and Table C2, respectively. As can be noted from these tables, our attempts to recover stochasticity (measurement errors) in travel time (which was free of measurement errors in the DGP) are innocuous. That is, the parameter recovery was not impacted. Specifically, the APB values are small. Further, the standard errors of the estimates were found to be similar to the those when we estimated the models without stochasticity in travel time. The trace of the covariance matrix of coefficients was 0.242 when imposing normal distributed stochasticity, which is close to the trace value of 0.246 we obtained from an estimation that did not impose stochasticity. [11]

[11] Clearly, we did not find much degradation in the consistency and the efficiency of the estimator when we assumed bus travel time variable to be stochastic when that variable was not stochastic (that is, when the DGP did not have stochasticity in the exogenous variable). Given this result, for forecasting purposes, if one uses the estimated model to predict mode choice for a given bus travel time value, our model that incorrectly assumes stochasticity (when such stochasticity is not present) does not do much harm. On the other hand, as discussed in Section 4, a modelling framework that ignores stochasticity in the exogenous variable (when stochasticity is present) leads to inconsistent estimation as well as an underestimation of standard errors. These issues will likely cause repercussions in forecasting, too.

Table C1 Simulation evaluation results where travel time is free of measurement errors but its distribution is estimated (normal travel time and coefficient on travel time)

	True value	***ICSV-RC* model results**			
		Mean	APB	FSSE	ASE
Bus travel time model					
Inverse speed θ_b: Location parameter	1.50	1.50	0.02	0.001	0.001
Inverse speed θ_b: Scale parameter	0.00	0.0003	--	0.002	0.002
SD of measurement error	0.95	0.95	0.11	0.007	0.007
Mean APB, FSSE, ASE	--	--	0.07	0.004	0.004
Mode choice model					
ASC for transit	-0.56	-0.55	2.35	0.046	0.066
ASC for walk	1.56	1.67	6.88	0.071	0.107
Travel time coefficient γ_{TT^*} - Location parameter	-1.00	-1.02	2.08	0.044	0.041
Travel time coefficient γ_{TT^*} -Scale parameter	0.19	0.20	3.31	0.014	0.015
Cost coefficient (β_C)	-0.25	-0.25	0.09	0.013	0.013
Mean of APB, FSSE, ASE	--	--	2.94	0.038	0.048

Table C2 Simulation evaluation results where travel time is free of measurement errors but its distribution is estimated (lognormal travel time and coefficient on travel time)

	True value	***ICSV-RC* model results**			
		Mean	APB	FSSE	ASE
Bus travel time model					
Inverse speed θ_b: Location parameter	0.49	0.49	0.001	0.0002	0.0003
Inverse speed θ_b: Scale parameter	0.00	0.0003	--	0.002	0.019
SD of measurement error	0.95	0.95	0.28	0.006	0.006
Mean APB, FSSE, ASE	--	--	0.141	0.003	0.003
Mode choice model					
ASC for transit	-0.56	-0.52	7.85	0.075	0.060
ASC for walk	1.56	1.28	17.92	0.098	0.117
Travel time coefficient γ_{TT^*} - Location parameter	-1.00	-0.98	1.54	0.039	0.039
Travel time coefficient γ_{TT^*} -Scale parameter	0.05	0.06	19.01	0.036	0.060
Cost coefficient (β_C)	-0.25	-0.26	2.55	0.011	0.009
Mean of APB, FSSE, ASE	--	--	9.77	0.052	0.057

References

Alvarez-Daziano, R., Bolduc, D., 2013. Incorporating pro-environmental preferences towards green automobile technologies through a Bayesian hybrid choice model. *Transportmetrica A: Transport Science*, 9(1), 74–106.

Batley, R. P., Toner, J. P., Knight, M. J., 2004. A mixed logit model of U.K. household demand for alternative-fuel vehicles. *International Journal of Transport Economics*, 31(1), 55–77.

Ben-Akiva, M., Walker, J., Bernardino, A. T., Gopinath, D. A., Morikawa, T., Polydoropoulou, A., 2002. Integration of choice and latent variable models. In Perpetual motion: Travel behaviour research opportunities and application challenges, 431-470. Pergamon, Amsterdam.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. Handbook of Transportation Science, Springer, 23, 5-33.

Bhat, C. R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, 35(7), 677-693.

Bhat, C. R., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B: Methodological*, 37(9), 837-855.

Bhat, C.R., Sardesai, R. (2006). The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B: Methodological*. 40(9), 709-730.

Bhat, C. R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B: Methodological*, 45(7), 923-939.

Bhat, C. R., Sidharthan, R., 2012. A new approach to specify and estimate non-normally mixed multinomial probit models. *Transportation Research Part B: Methodological*, 46(7), 817-833.

Bhat, C. R., Dubey, S. K., 2014. A new estimation approach to integrate latent psychological constructs in choice modeling. *Transportation Research Part B: Methodological*, 67(C), 68-85.

Bhatta, B. P., Larsen, O. I., 2011. Errors in variables in multinomial choice modeling: A simulation study applied to a multinomial logit model of travel mode choice. *Transport Policy*, 18(2), 326-335.

Biswas, M., Pinjari, A. R., Dubey, S. K., 2019, January. Travel Time Variability and Route Choice: An Integrated Modelling Framework. In 2019 11th International Conference on Communication Systems & Networks (COMSNETS), 737-742, IEEE.

Brownstone, D., Bunch, D. S., Train, K., 2000. Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research Part B: Methodological*, 34(5), 315-338.

Chen, A., Zhou, Z., Lam, W. H., 2011. Modeling stochastic perception error in the mean-excess traffic equilibrium model. *Transportation Research Part B: Methodological*, 45(10), 1619-1640.

Cherchi, E., Ortúzar, J. de D., 2008. Predicting best with mixed logit models: understanding some confounding effects. In: Inweldi, P.O. (Ed.), *Transportation Research Trends*. Nova Science Publishers, Inc., New York, pp. 215–235.

Conniffe, D., O'Neill, D., 2008. An efficient estimator for dealing with missing data on explanatory variables in a probit choice model. Available at SSRN 1284362.

Daly, A. J., Ortúzar, J. D., 1990. Forecasting and data aggregation: theory and practice. *Traffic Engineering and Control*, 31(12), 632-643.

De Palma, A., Lindsey, R., Picard, N., 2007. Congestion, Risk Aversion and the Value of Information. Working Paper 2, University of Alberta, Edmonton.

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

Díaz, F., Cantillo, V., Arellana, J., de Dios Ortúzar, J., 2015. Accounting for stochastic variables in discrete choice models. *Transportation Research Part B: Methodological*, 78(C), 222-237.

Dubey, S., Cats, O., Hoogendoorn, S., Bansal, P., 2022. A multinomial probit model with Choquet integral and attribute cut-offs. *Transportation Research Part B: Methodological*, 158(C), 140-163.

Durbin, J., 1954. Errors in variables. *Revue de l'institut International de Statistique*, 12(3), 23-32.

Fosgerau, M., & Fukuda, D. (2012). Valuing travel time variability: Characteristics of the travel time distribution on an urban road. *Transportation Research Part C: Emerging Technologies*, 24, 83-101.

Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological* 41(3), 363-378.

Gleser, L. J., 1981. Estimation in a multivariate "errors in variables" regression model: large sample results. *The Annals of Statistics*, 9 (1), 24-44.

Greene, W. H., Hensher, D. A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological*, 37(8), 681-698.

Heckman, J.J., 1976. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement*, 5(4), 475-492.

Heckman, J. J., 1979. Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161.

Hensher, D. A., Greene, W. H., 2003. The mixed logit model: the state of practice. *Transportation*, 30(2), 133-176.

Hess, S., Polak, J. W., 2005. Mixed logit modelling of airport choice in multi-airport regions. *Journal of Air Transport Management*, 11(2), 59-68.

James, J. Q. (2021). Citywide estimation of travel time distributions with Bayesian deep graph learning. *IEEE Transactions on Knowledge and Data Engineering,* 35(3), 2366-2378.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decisions under risk. *Econometrica,* 47(2), 263–291.

Li, Z., Hensher, D.A., 2011. Prospect theoretic contributions in understanding traveller behaviour: a review and some comments. *Transport Policy,* 31(1), 97–115.

Liu, X., Polak, J.W., 2007. Nonlinearity and the specification of attitudes towards risk in discrete choice models. In: Proceedings 86th Annual Meeting of the Transportation Research Board, Washington, DC.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. *Journal of applied Econometrics*, 15(5), 447-470.

Nirmale, S. K., Pinjari, A. R., 2023. Discrete choice models with multiplicative stochasticity in choice environment variables: Application to accommodating perception errors in driver behaviour models. *Transportation Research Part B: Methodological*, 170(C), 169-193.

Ortúzar, J. de D., Ivelic, A. M., 1987. Effects of using more accurately measured level-of-service variables on the specification and stability of mode choice models. In: Proceedings 15th PTRC Summer Annual Meeting, University of Bath, September 1987, England.

Ortúzar, J. de D., Willumsen, L. G., 2011. Modelling transport. John Wiley & Sons.

Patil, P. N., Dubey, S. K., Pinjari, A. R., Cherchi, E., Daziano, R., Bhat, C. R., 2017. Simulation evaluation of emerging estimation techniques for multinomial probit models. *Journal of choice modelling*, 23(C), 9-20.

Pinjari, A.R., Kamali, M., Luong, T., Ozkul, S., 2015. *GPS Data for Truck-Route Choice Analysis of Port Everglades Petroleum Commodity Flow*. Report BDV25-977-17. Florida Department of Transportation.

Polak, J.W., Hess, S., Liu, X., 2008. Characterizing heterogeneity in attitudes to risk in expected utility models of mode and departure time choice. In: Proceedings 87th Annual Meeting of the Transportation Research Board, Washington, DC.

Rahmani, M., Jenelius, E., & Koutsopoulos, H. N. (2015). Non-parametric estimation of route travel time distributions from low-frequency floating car data. *Transportation Research Part C: Emerging Technologies,* 58, 343-362.

Rasouli, S., Timmermans, H. (2014). Applications of theories and models of choice and decision-making under conditions of uncertainty in travel behavior research. *Travel Behaviour and Society*, 1(3), 79-90.

Revelt, D., Train, K., 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. *Review of economics and statistics*, 80(4), 647-657.

Rieser-Schüssler, N., Balmer, M., Axhausen, K.W., 2013. Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science* 9(9), 825-845.

Rubin, D., 1987. Multiple Imputation for Nonresponse in Surveys. Wiley, New York.

Sanko, N., Hess, S., Dumont, J., Daly, A., 2014. Contrasting imputation with a latent variable approach to dealing with missing income in choice models. *Journal of Choice Modelling*, 12(C), 47-57.

Senbil, M., Kitamura, R., 2004. Reference points in commuter departure time choice: a prospect theoretic test of alternative decision frames. *Journal of Intelligent Transportation Systems*, 8(1), 19–31.

Srinivasan, K. K., Prakash, A. A., Seshadri, R., 2014. Finding most reliable paths on networks with correlated and shifted log–normal travel times. *Transportation Research Part B: Methodological*, 66(C), 110-128.

Stefanski, L. A., Carroll, R. J., 1985. Covariate measurement error in logistic regression. *The Annals of Statistics*, 13(4), 1335-1351.

Steinmetz, S.S., Brownstone, D., 2005. Estimating commuters' "value of time" with noisy data: a multiple imputation approach. *Transportation Research Part B: Methodological*, 39(3), 865-889.

Swait, J.D., Bernardino, A., 2000. Distinguishing taste variation from error structure in discrete choice data. *Transportation Research Part B: Methodological*, 34(1), 1–15.

Swait, J., 2022. Distribution-free estimation of individual parameter logit (IPL) models using combined evolutionary and optimization algorithms. *Journal of Choice Modelling*, 47(C), 100396.

Tahlyan, D., Luong, T., Pinjari, A., Ozkul, S., 2017. Development and Analysis of Truck Route Choice Data for the Tampa Bay Region using GPS Data. Report BDK25-730-3. Florida Department of Transportation.

Tahlyan, D, Pinjari, A., 2020. Performance evaluation of choice set generation algorithms for analyzing truck route choice: insights from spatial aggregation for the breadth first search link elimination (BFS-LE) algorithm. *Transportmetrica A: Transport Science*, 16(3). 1030-1061.

Thakur, A., Pinjari, A.R., Zanjani, A.B., Short, J., Mysore, V., Tabatabaee, S.F., 2015. Development of Algorithms to Convert Large Streams of Truck GPS Data into Truck Trips. *Transportation Research Record: Journal of the Transportation Research Board* 2529(1), 66-73.

Train, K., 1978. A validation test of a disaggregate mode choice model. *Transportation Research*, 12(3), 167-174.

Train, K. E. (1998). Recreation demand models with taste differences over people. *Land Economics,* 74(2), 230-239.

Tversky, A., Kahneman, D., 1992. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty,* 5(4), 297–323.

Varela, J.M.L., Daly, A., Börjesson, M. (2018) Quantifying errors in travel time and cost by latent variables, *Transportation Research B: Methodological*, 117(PA), 520-541.

Varotto, S. F., Glerum, A., Stathopoulos, A., Bierlaire, M., Longo, G., 2017. Mitigating the impact of errors in travel time reporting on mode choice modelling. *Journal of Transport Geography*, 62(C), 236-246.

Vij, A., Walker, J. L., 2016. How, when and why integrated choice and latent variable models are latently useful. *Transportation Research Part B: Methodological*, 90(C), 192-217.

Von Neumann, J., Morgenstern, O., 1947. Theory of Games and Economic Behavior, second ed. Princeton University Press, Princeton, NJ.

Walker, J., Li, J., Srinivasan, S., Bolduc, D., 2010. Travel demand models in the developing world: Correcting for measurement errors. *Transportation Letters*, 2(4), 231-243.

Zhang, K., Jia, N., Zheng, L., & Liu, Z. (2019). A novel generative adversarial network for estimation of trip travel time distribution with trajectory data. *Transportation Research Part C: Emerging Technologies*, 108, 223-244.