Discrete Choice Models with Multiplicative Stochasticity in Choice Environment Variables: Application to Accommodating Perception Errors in Driver Behaviour Models

Sangram Krishna Nirmale [a]

Email: nirmalek@iisc.ac.in

Abdul Rawoof Pinjari [a,b]

Email: abdul@iisc.ac.in

[a] Department of Civil Engineering, Indian Institute of Science, Bengaluru, 560012, India.

[b] Centre for infrastructure, Sustainable Transportation & Urban Planning (CiSTUP), Indian Institute of Science (IISc), Bengaluru, 560012, India.

ABSTRACT

This paper presents a mixed multinomial logit-based discrete choice modelling framework to accommodate decision-makers' errors in perceiving choice environment variables that do not vary across choice alternatives. An analysis is undertaken to evaluate two different ways of specifying errors in the choice environment variables in discrete choice models – (a) the additive specification and (b) the multiplicative specification. Between these two approaches, the multiplicative error specification is consistent with psychophysical theories of human perception of physical quantities in that the variability in perception tends to be greater for quantities of greater magnitude. Further, it is shown that models with an additive error specification run into parameter (un)identifiability problems if the analyst attempts to accommodate errors in several variables. In contrast, models with multiplicative errors in variables allow separate identification of stochasticity in as many variables as needed, as long as those variables have a significant influence on the choice outcome.

The usefulness of the proposed framework with multiplicative errors is demonstrated through simulation experiments and an empirical application for analysing driver behaviour while considering drivers' errors in perceiving traffic environment variables. The empirical analysis is carried out using space-time trajectories of vehicles from a heterogeneous, disorderly (HD) traffic stream in Chennai, India. Results suggest that the proposed model, with power lognormal distributed multiplicative errors in traffic environment variables, outperformed the typically used mixed logit models with random coefficients (uncorrelated and correlated) or error components. Further, allowing for perception errors in traffic environment variables was found to be more important than allowing unobserved heterogeneity in the drivers' sensitivity to those variables. In addition, the empirical model offers interesting insights into the extent of variability due to perception errors in different traffic environment variables.

Keywords: *mixed logit, errors in variables, perception errors, parameter identification, driver behaviour, heterogeneous and disorderly traffic*

1. INTRODUCTION

Random utility maximization (RUM) based discrete choice models involve utility functions that are typically specified as functions of observed variables describing choice alternative attributes, decision-maker characteristics, and choice environment variables. In addition, the utility functions include random error terms to recognize differences between the systematic utility components characterized by the analyst and the utility perceived by the decision-maker. As discussed in Manski (1977), the random error terms include, for example, omitted attributes that have an influence on the decision-maker's utility, unobserved taste variations, measurement errors in the variables included in systematic utility components, and other errors in the utility specification. In addition, even if the analyst had access to accurate measurements, the random error terms would include perception errors of the decision-makers.

Some of the above reasons for including random error terms, such as taste variations, may be addressed by treating the parameters of the systematic utility function as random. A large stream of literature exists on random coefficients in choice models (Cardell and Dunbar, 1980; McFadden and Train, 2000). However, several other reasons for stochasticity in utility functions, such as measurement and/or perception errors for variables included in the systematic utility functions, warrant the treatment of those variables as stochastic. For example, using aggregate, zone-to-zone measurements instead of point-to-point measurements (Train, 1978; Daly and Ortuzar, 1990) or assuming free-flow travel times can introduce errors in the travel time variables used to explain many travel choices. Spatial aggregation can introduce errors in spatial variables used in location choice models (Daly and Ortuzar, 1990; Hellerstein, 2005). In some situations, noisy data might be a reason for errors in variables (Steinmetz and Brownstone, 2005; Walker et al., 2010; Bhatta and Larsen, 2011). In another example, in models of driver behaviour in traffic streams, traffic environment variables such as space gaps and relative speeds are typically treated as deterministic. However, drivers' perceptions of these variables might be different from the analysts' measurements typically included as explanatory variables in the models. All these reasons warrant the need to accommodate uncertainty in the explanatory variables used in models of choice behaviour.

1.1. Choice Models with *Errors in Variables* (EIV)

The literature on choice models with *errors in variables* (EIV) is relatively small compared to that on choice models with random parameters. As pointed out by McFadden (1984) and recently brought to attention by Díaz et al. (2015), the EIV issue poses important yet not fully resolved problems for choice modelling. This issue has received greater attention in the econometric literature, particularly in the form of EIV in linear regression models (Fuller, 2009; Greene, 2018) and to some extent in non-linear models (Wansbeek and Meijer, 2000; Carroll et al., 2006). As such, there is a consensus in the econometric literature on non-linear models that EIV can potentially result in biased parameter estimates (due to endogeneity) not only for the variables with errors but also for other variables in the model (Greene, 2018). This is because the endogeneity caused by EIV typically affects the estimation of all parameters in non-linear models (Wooldridge, 2012). The issue of endogeneity arises when the EIV are correlated with one or more explanatory variables in the model. However, in RUM-based discrete choice models, even if the EIV are not correlated with the explanatory variables, ignoring EIV would lead to an inflation of variance of the kernel error terms, thereby causing bias toward zero for the parameter estimates (because parameter estimates in discrete choice models are confounded by the scale of the kernel error terms). Several studies in the choice modelling literature discuss and/or demonstrate that ignoring stochasticity due to EIV, when present, can potentially lead to biased estimation and distorted inferences (Yatchew and Griliches, 1985; Hellerstein, 2005; Carroll et al., 2006; Bhatta and Larsen, 2011; Díaz et al., 2015), incorrect marginal rates of substitution (Ortúzar and Ivelic, 1987; Bhatta and Larsen, 2011), and erroneous forecasts (Train, 1978, 2009).

To address the EIV problem in choice models, a stream of studies in the biometrics field (for example, Carroll et al., 1984; Stefanski and Carroll, 1985) propose bias-adjusted estimators for binary choice models and a few studies in the economics field (Kao and Schnell, 1987) do the same for multinomial logit models. In another study, Steinmetz and Brownstone (2005) use an imputation method (Rubin, 1987) to correct measurement errors in network data, such as travel times, when accurate measurements are available only for a sub-sample of observations.

In another widely used approach to address the EIV problem, the variables under consideration are treated as latent. Available measurements of the variables are used to inform the distribution of the latent variables through a measurement equation. The latent variable, in turn, enters the utility function of the choice model. The measurement equation and the choice model are estimated jointly in an integrated choice and latent variable (ICLV) framework (Bolduc and Alvarez-Daziano, 2010; Walker et al., 2010; Sanko et al., 2014; Varotto et al., 2017; Biswas et al., 2019). In most such ICLV studies, separate structural equations are specified for the latent variables under consideration, where the latent variables are expressed as functions of exogeneous variables. For example, income may be specified as a function of sociodemographic characteristics (Sanko et al., 2014), and route-level travel time may be specified as a function of route structure attributes. Doing so, however, is not always possible, especially when it is not easy to find exogenous variables to explain the latent variable. In such situations, the latent variable is expressed as a sum of available measurements and a random error term to recognize the error in the measurement. This approach is used to account for EIV in a multinomial choice model by Hellerstein (2005) and Díaz et al. (2015). In both papers, the authors deal with errors in alternative attributes – location-specific attributes in a location choice model by Hellerstein (2005) and travel time variables in a mode choice model by Díaz et al. (2015). Furthermore, in both papers, the EIV specification is converted into an error components specification where the EIV in all variables of interest are combined into one error component for each choice alternative. The resulting model, assuming IID Gumbel kernel error terms, is the familiar mixed multinomial logit model with a heteroscedastic structure. A downside of this approach is that one cannot separately estimate error components for each explanatory variable with errors because, for each choice alternative, only a single variance term can be estimated (while normalizing the variance for one alternative). Besides, it is difficult to estimate separate error component parameters for each choice alternative in situations with large choice sets.

1.2. Gaps in Literature

There are three prominent gaps in most of the above-discussed literature on choice models with EIV. First, most of the above-discussed studies focus on errors in choice alternative attributes that vary across alternatives, such as travel times in mode choice or route choice models. Few studies focus on errors in choice environment variables that do not vary across choice alternatives. However, several choice environment variables that do not vary across choice alternatives, such as drivers' perceptions of their traffic environment in driver behaviour models, can potentially be associated with errors. And there is one important difference between the errors in these two types of variables. Errors in choice environment variables that do not vary across alternatives must be represented by the same probabilistic distribution across all choice alternatives. This is because the decision-makers' errors in perceiving a choice environment variable do not vary across choice alternatives. On the other hand, the distributions for errors in alternative-specific attributes are typically different for different choice alternatives. For example, variability in travel times of bus transit can potentially be higher than that of metro transit. Therefore, the specification of errors in choice environment variables cannot be the same as that for alternative-specific attributes.

Second, most of the above-discussed literature is in the context of accommodating measurement errors. However, in several situations, the decision-maker's errors in perceptions of physical quantities – such as time duration, distance, and speed – might be more prevalent than the analyst's errors in measuring the true values of those quantities. In such cases it becomes important to recognize the errors in decision-maker's perception of the variables under consideration.[1]

To be sure, there is a stream of literature that accounts for decision-maker's perception errors in choice models. For example, the stochastic user equilibrium model of route choice (Daganzo and Sheffi, 1977) is based on stochasticity in utility functions due to perception errors in route-level travel times. Further, route choice applications of discrete choice models with multiplicative random utility terms (Fosgerau and Bierlaire, 2009) also motivate perception errors as a reason for multiplicative error terms. Another study on the value of time estimation by Hess et al. (2017) motivates the use of multiplicative errors for the utility function to capture context

[1] One might suggest that that the decision-maker's perception errors can be treated as the analyst's errors in measuring the decision-maker's perceptions. However, it is useful to treat decision-makers' errors in perceiving physical quantities separately from the analyst's measurement errors. In this context, note that the analyst can make two types of measurements – (1) measurement of the true value of the physical quantity and (2) measurement of the decision-maker's perceived value of the physical quantity. However, most often, empirical studies have access to analyst's measurements of the true value (perhaps with some error) than the analyst's measurements of the decision-maker's perceptions. In many contexts (e.g., driver behaviour), it is much more difficult to measure decision-makers' perceived values than to measure the true value of a physical quantity. Even in contexts such as mode choice, the analyst may have access to travelers' perceptions of the attributes (e.g., reported travel times) of only their chosen modes. It is not easy to elicit travelers' perceptions of the attributes of a mode they did not choose. Therefore, we use the term *perception error* to represent the gap between the true value and the decision-maker's perceived value of a variable. The term *measurement error* may be used to represent the gap between the true value and the analyst's measurement of true value of the variable.

Further, as will be discussed in Section 3.2, theories of human perception may be invoked to guide the approach to specifying stochasticity due to perception errors (i.e., the gaps between true and perceived values). However, no theoretical guidance is available if one treats the gaps between the analyst's measurement of the true values and the decision-maker's perceived values as measurement errors (recall that the analyst was not even trying to measure the perception). If both sources of error – decision-maker's perception and analyst's measurement of the truth – are prevalent, it is better to represent both these sources separately and bring to bear theory and data to inform both sources of stochasticity than to combine them and then try to characterize the resulting stochasticity. Finally, in contexts such as driving behaviour, which is an important field of study, the decision-maker's perception errors are likely to be more prevalent than the analyst's measurement errors. This is because drivers perceive and estimate the characteristics of their choice environment in real-time, whereas the analysts measure the same characteristics offline. Since much care is taken in deriving the measurements from data sources such as traffic videos, it is defensible to assume that the variability in analyst's errors in measuring the true values is negligible than that due to drivers' errors in perceiving the true values.

effects, such as greater variability for longer trips. To the authors' knowledge, most of these studies focus on perception errors in alternative attributes or context effects on the overall utility function, not on specific choice environment variables that do not vary across choice alternatives but are included with alternative-specific coefficients in the utility functions.

Third, most literature on accommodating EIV does so through an additive specification of errors in the variables, where the error term specific to a variable is added to the measurement of that variable. However, as will be discussed in Section 3.2, psychophysical theories of human perception of physical quantities motivate the need for using multiplicative errors for capturing perception error. In such a specification, the error term specific to a variable is multiplied to the measurement of that variable. As a result, the variability due to error in perception increases with the magnitude of the quantity being perceived, a pattern that is not straightforward to capture using the additive EIV specification. Further, as will be shown in Section 3.3, the additive approach to specifying EIV does not help in identifying variability due to errors in variables that do not vary across alternatives (if there are several such variables with perception errors). Only a few studies explore the multiplicative error specification on variables in the utility function. For example, Varela et al. (2018) explore both additive and multiplicative errors in latent variables to account for measurement errors in travel times and travel costs. However, most such studies do not delve into attributes that do not vary across alternatives nor focus on the perception errors of travellers.

1.3. Current Study

In this study, we present a discrete choice modelling framework to accommodate stochasticity in choice environment variables that do not vary across choice alternatives. The stochasticity may arise due to various reasons – decision-makers' errors in perceiving the choice environment, analyst's errors in measuring such variables, or inherent stochasticity of the variables. In this paper, we focus on the decision-makers' errors in perception as the primary source of stochasticity. The model structure takes the form of a mixed multinomial logit (ML) model where the choice environment variables under consideration are specified as stochastic. To operationalize this framework, we evaluate two different ways of specifying errors in choice environment variables in discrete choice models – (a) the additive EIV specification (error term specific to a variable is added to the measurement of that variable) and (b) the multiplicative EIV specification (error term specific to a variable is multiplied to the measurement of that variable). Using the multiplicative

EIV specification, it is easy to accommodate that quantities of larger (smaller) magnitude are perceived with greater (smaller) variability. Further, we show that models with an additive error specification run into parameter (un)identifiability issues if the analyst attempts to recognize errors in more choice environment variables than the number of choice alternatives minus one. On the other hand, models with multiplicative errors are not saddled with such identification problems. In fact, in theory, and if data allows, one can attempt to recover multiplicative stochasticity separately for as many choice environment variables as needed.

We also discuss the possibility of confounding between the proposed multiplicative EIV specification on choice environment variables and correlated random coefficients on the same variables. In this context, we show that a correlated random coefficients model is a more general specification that subsumes our proposed model with multiplicative EIV as a special case. Despite such confounding, we demonstrate that the estimation of such a general specification is not possible (due to parameter unidentifiability) if the source of stochasticity is predominantly multiplicative errors in the choice environment variables, as opposed to random coefficients on those variables. In such situations, the analyst should estimate the proposed multiplicative EIV model as opposed to the more general, correlated random coefficients model.

The proposed choice model with multiplicative errors on explanatory variables is applied to accommodate drivers' perception errors in a multi-stimuli-based model of driver behaviour in *heterogeneous, disorderly* (HD) traffic streams using space-time trajectories of vehicles from an arterial road in Chennai, India. Specifically, a subject vehicle's (SV) driver behaviour in the traffic stream is represented as a choice from a set of discrete alternatives – accelerate, decelerate, or maintain the same speed – at any given time. Variables used to represent the driver's perception of the traffic environment, such as space gaps and relative speeds with respect to other vehicles, are considered stochastic to recognize the errors drivers make in perceiving those quantities.

Before proceeding with the empirical analysis, simulation experiments are carried out for the afore-mentioned choice context to evaluate the parameter recovery of the proposed model using the maximum simulated likelihood (MSL) estimation method. In addition to the proposed ML model with multiplicative perception errors (i.e., multiplicative EIV), we explore the efficacy of alternative ML models with random coefficients (instead of stochastic variables) and those with error components on the same simulated data. In doing so, we demonstrate that the estimation of a general, correlated random coefficients specification is not possible if the predominant source of stochasticity is multiplicative errors in the choice environment variables, as opposed to random coefficients on those variables. Further, in some empirical contexts, since the analyst may not know *apriori* whether to focus on stochasticity in decision-makers' response to choice environment variables, or their errors in perceiving those variables, or both, we conduct additional simulation experiments to develop guidelines for selecting a model structure and interpreting it.

In the empirical analysis, we explore alternative distributions for specifying multiplicative errors on choice environment variables. In addition, using both the estimation dataset and a validation dataset, we assess the importance of accommodating multiplicative perception errors separately for each choice environment variable. The empirical analysis also offers insights into the magnitudes of variability due to perception errors in different traffic environment variables.

In the rest of this paper, Section 2 reviews the literature on driver behaviour models that consider perception errors and highlights how our empirical study contributes to this literature. Section 3 describes the proposed model structure, along with an analysis to identify an appropriate specification to accommodate perception errors in choice environment variables in discrete choice models of driver behaviour. In Section 4, details of the vehicle trajectory dataset used in this paper are presented. Section 5 presents the simulation experiments and findings from the experiments. Section 6 presents the empirical analysis and discusses the empirical findings. Section 7 summarizes the paper and directions for future research.

2. DRIVER BEHAVIOUR MODELS WITH PERCEPTION ERRORS

For several decades, the typical car-following framework has been used to model driver behaviour, where the driver's acceleration/deceleration actions are modelled as a response to stimuli from a lead vehicle ahead of the driver's vehicle. In addition to vehicle kinematics and traffic environment variables, the literature abounds with studies highlighting the importance of human factors in these models. The human factors include, for example, drivers' socio-demographics, physiological factors, personality traits, and driving skills and desires (Hamdar, 2012; Treiber and Kesting, 2013; Saifuzzaman and Zheng, 2014; Sharma et al., 2018). Here, we focus on driver behaviour models that consider errors in drivers' perception of their traffic environment.

Consideration of drivers’ perception errors has long been recognized as important for improving the realism of driver behaviour models. For example, Gray and Regan (1998) demonstrate that the driver’s perceptions of ‘distances to', ‘velocities of’, and ‘accelerations of’ other objects are not exact. Wiedemann (1974) recognizes that drivers cannot perceive stimuli below a minimum threshold value and proposes a psychophysical driver behaviour model with perception thresholds. Hoogendoorn et al. (2011) use this model and present a stochastic car-following model whose thresholds are determined empirically from vehicle trajectory data. Further, Kikuchi and Chakroborty (1992) use fuzzy sets to represent the approximate nature of drivers’ decision processes. Treiber et al. (2006) use the Wiener stochastic processes to describe drivers’ perception errors for relative positions, speeds, and speed differences. This study concludes that errors in estimation (perception error) are influential on driver behaviour and affect the performance and stability of the vehicular traffic stream. Van Lint et al. (2017) also use the Wiener process to model perception error in their model for integrated analysis of lane-changing and car-following behaviour. Yang and Peng (2010) propose an *errorable* car-following model that considers human reaction delays, distraction, and perception limitations. Using a similar line of thought, Bevrani and Chung (2012) improve Gipps’ (1981) model to accommodate human imperfection in perceiving and processing information and executing actions.

In another stream of literature, random utility maximization-based discrete choice models have been used to analyse various aspects of driver behaviour, including acceleration/deceleration decisions, lane-changing behaviour, lateral position choices, and gap-acceptance (Ahmed, 1999; Toledo, 2003; Choudhury, 2007). Additionally, latent variables have been used in such models to represent variables unobserved to the analyst, such as latent plans, latent intents, latent leaders, and reaction times (Choudhury, 2007; Koutsopoulos and Farah, 2012; Choudhury and Islam, 2016), but not to represent drivers’ errors in perceiving their traffic environment. All these studies consider single-leader car-following behaviour. Further, most utility-based driver behaviour models do not consider drivers’ perception errors (except, for example, Hamdar et al., 2015).

Many studies discussed above do not demonstrate any evidence of perception error in empirical data. They formulate stylized models and conduct simulation and/or numerical experiments to understand driver behaviour in the presence of perception error; therefore, the empirical evidence available is very limited in this context. Further, current literature does not recognize that the level of errors in perception might be different for different variables and different surrounding vehicles in the traffic environment. Besides, most studies that consider drivers' perception error are in the context of a single-leader car-following setting in homogeneous traffic conditions. Given the lack of lane discipline in heterogeneous traffic streams observed in many countries, multiple vehicles around a vehicle might influence its drivers' behaviour. In such traffic environments, drivers need to perceive and process multiple sources of stimuli for making their manoeuvering decisions. Therefore, errors are likely prevalent in their perception of traffic environment variables such as distances and relative speeds.

This study attempts to fill the above-mentioned gaps by considering perception errors in a multi-stimuli-based driver behaviour model that considers different levels of perception errors in different traffic environment variables with respect to different vehicles around the subject vehicles. The proposed framework utilizes stochastic variables to recognize that the drivers' perceptions of traffic environment variables are likely to be associated with errors. Doing so allows the analyst to (a) empirically assess and compare the extent of stochasticity (or variance) due to drivers' perception errors for different variables of influence on driver behaviour and (b) examine the importance of accounting for such stochasticity on driver behaviour.

3. METHODOLOGY

3.1. Model Structure

Let q and i be the indices representing subject vehicles and their discrete manoeuvring choice alternatives ($a =$ accelerate, $d =$ decelerate, $s =$ maintain same speed), respectively, and let x^*_{qk} denote the driver's perceived value of the k^{th} traffic environment variable, whose measured value by the analyst is x_{qk}. Stack all the traffic environment variables x^*_{qk} perceived by a driver of vehicle q into a vector x^*_q. The driver-perceived values x^*_q are treated as stochastic variables that are known only up to an assumed distribution. The parameters (θ) of the distribution $f(x^*_q;\theta \mid x_q)$ of such stochastic variables may be identified using the analyst's measurements (x_q) of those variables and the drivers' behaviour. In this context, it is assumed that the measurements (x_q) typically obtained from observed vehicle trajectory datasets are free of errors (see Footnote 1).

Consider the following utility specification for each of the discrete manoeuvring choice alternatives faced by the driver of the vehicle q:

$$U_{qi} = \beta_{i0} + \sum_{k=1}^{K} \beta_{ik} x_{qk}^{*} + \xi_{qi} \tag{1}$$

In this equation, U_{qi} is the utility of manoeuvring alternative i for the driver of vehicle q. β_{i0} is the constant specific to i, β_{ik} $(k = 1, 2, ..., K)$ are unknown parameters to be estimated and refer to the influence of the corresponding perceived variables x_{qk}^{*} ($k = 1,2,...,K$) on the preference for manoeuvring alternative i, and ξ_{qi} is an error term assumed to be independently and identically (IID) Gumbel distributed. Following the random utility maximization theory, the driver of the subject vehicle q is assumed to choose a manoeuvring alternative i if $U_{qi} > U_{qj}$ $\forall i \neq j$.

The conditional likelihood $L_{qi}\left(\beta, x_{q}^{*}\right)$ that the driver of vehicle q makes a manoeuvring choice i given the traffic environment variable values x_{q}^{*} is the following logit function:

$$L_{qi}\left(\beta, x_{q}^{*}\right) = \frac{\exp\left(\beta_{i0} + \sum_{k=1}^{K} \beta_{ik} x_{qk}^{*}\right)}{\sum_{j=a,d,s} \exp\left(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} x_{qk}^{*}\right)} \tag{2}$$

In this equation, β in the left side of the equation is a vector of parameters obtained by stacking the β_{i0} and β_{ik} parameters of all choice alternatives. Similarly, x_{q}^{*} is obtained by stacking all x_{qk}^{*} variables $(k = 1, 2, ..., K)$. Assuming a distribution $f\left(x_{q}^{*}; \theta\right)$ for x_{q}^{*} and integrating the conditional likelihood over the distribution of x_{q}^{*} results in the following unconditional likelihood expression:

$$L_{qi}(\beta, \theta) = \int_{x_{q}^{*}} L_{qi}\left(\beta, x_{q}^{*}\right) f\left(x_{q}^{*}; \theta\right) dx_{q}^{*} \tag{3}$$

Assuming independence across all observations (q), the likelihood for the entire data is a product of the likelihoods of observed choices across all observations. The unknown parameter vector (β, θ) can be estimated using the maximum simulated likelihood (MSL) estimation routine. Appendix A provides details on the estimation of the proposed model, including its simulated likelihood function and expressions for the gradients of the simulated likelihood function.

The likelihood expression in Eq. (3) is a mixed logit likelihood expression. However, unlike the typical mixed logit models where the coefficients (β_{ik}) are random, the above model assumes the explanatory variables (x^*_{qk}) as random while keeping deterministic coefficients. It is worth noting here that the stochasticity in explanatory variables (x^*_{qk}) can potentially be confounded with stochasticity in coefficients (β_{ik}) if the random coefficients on a stochastic variable are correlated across different choice alternatives. This issue is discussed in detail in Section 3.4. Despite such confounding, simulation experiments in Section 5 help us identify when a model with stochastic choice environment variables is more suitable than a correlated random coefficients model.

3.2. Specification of the Stochastic Variables (x^*_{qk})

The most common approach to specifying errors in variables assumes that the magnitude of error is independent of the observed/measured value. Under this assumption, the perceived value by the driver of the subject vehicle q for the k^{th} variable may be expressed as:

$$x^*_{qk} = x_{qk} + \eta_{qk} \tag{4}$$

where, η_{qk} is a normally distributed error component with an expected value of zero and standard deviation σ_k (other distributional assumptions may also be explored). Normalizing the mean of the error to zero assumes zero bias in perception (with respect to the measurement)[2]. However, this normalization is not sufficient to identify the model with additive error specification for choice environment variables that do not vary across choice alternatives. More on this in Section 3.3.

An alternative to the classical additive error structure is the multiplicative structure, where the perceived value (x^*_{qk}) of a choice environment variable is expressed as a product of the measured value (x_{qk}) and a random error term (τ_{qk}), as below:

$$x^*_{qk} = x_{qk} . \tau_{qk} \qquad (5)^3$$

Assuming no bias in perception with respect to measurement (i.e., no difference in the expected value of x^*_{qk} and x_{qk}), the random error τ_{qk} should be specified to have an expected value equal

[2] Our assumption of zero bias with respect to measurement is made for the convenience of identification in the absence of additional information to inform bias in perception. However, there is a body of psychophysics literature on how human perception of time and other physical quantities is proportional to the magnitude of the quantity being perceived and that the bias in perception can be incorporated in the proportionality constant (Fechner et al., 1966). The issue of bias in perception is an avenue for further research.
[3] In another line of literature, multiplicate specification is used for the kernel error terms to develop alternative discrete choice models (see Castillo et al., 2008; Fosgerau and Bierlaire, 2009; Chikaraishi and Nakayama, 2016; Ojeda-Cabral et al., 2016) In this study, we stay within the class of additive-RUM models where the kernel error term is additively specified.

to one, i.e. $E[\tau_{qk}]=1$. This normalization helps in identification as well. In this paper, we label the proposed choice models with multiplicative perception errors in choice environment variables as ML-ME models (for mixed multinomial logit models with multiplicative errors).

A behavioural reason for specifying perception errors in the multiplicative form is that the errors humans make in perceiving physical quantities, such as distances, time duration, and speeds, depend on the magnitude of the quantity being perceived (Fechner et al., 1966). This observation is consistent with the intuition that larger (smaller) values of the quantity being perceived have larger (smaller) variability in perception. In the context of human perception of time duration, for example, Allan (2001) utilizes Weber's law from the field of psychophysics to state that the standard deviation of human perception of time duration is directly proportional to the mean of the perceived duration. Some of this literature is discussed in detail in a recent paper by Chakroborty et al. (2021), who state that "*...multiplicative errors are a natural choice while handling random variability in perceptions of not only time but also of other physical quantities.*" This is because multiplicative errors allow naturally for the variability to be larger for quantities of larger magnitude. Therefore, in situations where the analyst believes the gap between analyst-measured and decision-maker's perceived quantities is primarily due to the decision-maker's perception errors, a multiplicative error specification may be preferred. Besides, physical quantities such as space gaps widely used in driver behaviour models should not take negative values. While relative quantities such as relative speeds can be negative, it is reasonable to assume that people do not perceive a positive relative speed as negative or vice versa. Therefore, the distributions used to represent user perceptions of such quantities should not flip the sign of observed values. Multiplicative errors using distributions with support on the right half of the real line easily satisfy the above requirements while also allowing both larger values (overestimation) and smaller values (underestimation) than the observed values.

3.3. Identification of Stochasticity in Choice Environment Variables

Consider a driver's choice occasion with three alternatives – accelerate (a), decelerate (d), and maintain same speed (s) – with the corresponding utility functions denoted as U_{qa}, U_{qd} and U_{qs}, respectively, and three traffic environment variables that do not vary across alternatives (x_{q1}, x_{q2}, and x_{q3}) entering the utility functions. Specifically, consider the following utility structure:

$$
\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}(x^*_{q1}) + \beta_{a2}(x^*_{q2}) + \beta_{a3}(x^*_{q3}) + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}(x^*_{q1}) + \beta_{d2}(x^*_{q2}) + \beta_{d3}(x^*_{q3}) + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned} \tag{6}
$$

Note that, in the above specification, the traffic environment variables enter the utility functions with alternative-specific coefficients. They are not interacted with any alternative attributes because the choice alternatives do not have their own attributes that vary across alternatives. Such choice contexts without alternative attributes in the specification are common in driver behaviour analysis, activity-type choice analysis, and many other contexts. On the other hand, in choice contexts where alternative attributes are also present in the utility functions, the choice environment variables can enter the utility functions in one or both of the following two ways: (1) without interactions with alternative attributes, (2) through interactions with alternative attributes. The discussion in the current section is specific to the case when the choice environment variables are not interacted with alternative attributes.

3.3.1. Identification for Additive Specification of Error in Choice Environment Variables

Employing the additive error specification of Eq. (4) for choice environment variables, the utility structure in Eq. (6) may be written as:

$$
\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}x_{q1} + \beta_{a2}x_{q2} + \beta_{a3}x_{q3} + \beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}x_{q1} + \beta_{d2}x_{q2} + \beta_{d3}x_{q3} + \beta_{d1}\eta_{q1} + \beta_{d2}\eta_{q2} + \beta_{d3}\eta_{q3} + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned} \tag{7}
$$

Let the random components of the above utility functions be written as:

$$
\begin{aligned}
\varepsilon_{qa} &= \beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa}, \\
\varepsilon_{qd} &= \beta_{d1}\eta_{q1} + \beta_{d2}\eta_{q2} + \beta_{d3}\eta_{q3} + \xi_{qd}, \text{ and} \\
\varepsilon_{qs} &= \xi_{qs}
\end{aligned} \tag{8}
$$

Without loss of generality, assume that: (a) the additive perception error terms η_{qk} are normally distributed with zero mean and variance $\sigma^2_{\eta_k}$, and (b) the kernel error terms ξ_{qj} are IID Gumbel distributed with zero mean and scale parameter g_ξ, with the corresponding variance as $\sigma^2_{\xi j} = \pi^2 / 6g^2_\xi$. The variance-covariance matrix of the random utility terms $(\varepsilon_{qa}, \varepsilon_{qd}, \varepsilon_{qs})$ may be derived as below (Nirmale, 2022):

$$\Omega = \begin{bmatrix} \beta_{a1}^2\sigma_{\eta_1}^2 + \beta_{a2}^2\sigma_{\eta_2}^2 + \beta_{a3}^2\sigma_{\eta_3}^2 + \sigma_\xi^2 & \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 & 0 \\ \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 & \beta_{d1}^2\sigma_{\eta_1}^2 + \beta_{d2}^2\sigma_{\eta_2}^2 + \beta_{d3}^2\sigma_{\eta_3}^2 + \sigma_\xi^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{bmatrix} \quad (9)$$

The corresponding variance-covariance matrix for error differences (with respect to the base alternative, maintain same speed) is:

$$\Omega_\Delta = \begin{bmatrix} \beta_{a1}^2\sigma_{\eta_1}^2 + \beta_{a2}^2\sigma_{\eta_2}^2 + \beta_{a3}^2\sigma_{\eta_3}^2 + 2\sigma_\xi^2 & \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 + \sigma_\xi^2 \\ \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 + \sigma_\xi^2 & \beta_{d1}^2\sigma_{\eta_1}^2 + \beta_{d2}^2\sigma_{\eta_2}^2 + \beta_{d3}^2\sigma_{\eta_3}^2 + 2\sigma_\xi^2 \end{bmatrix} \quad (10)$$

Observe from the above variance-covariance matrix that its elements do not vary across individuals. For examining the identifiability of a model with such a variance-covariance matrix, both the order condition (necessary) and the rank condition (sufficient) must be employed (Bunch, 1991; Walker, 2001). The order condition states the maximum number of parameters that can be estimated, which depends on the number of alternatives in the choice set. The rank condition provides the actual number of parameters that can be estimated based on the postulated covariance structure. Finally, the positive definiteness of the covariance matrix must be verified to determine a valid normalization such that the hypothesized model's true structure is maintained when normalization restrictions are applied.

When discussing the order condition, it is useful to separate the covariance matrix (Ω) into two portions – a *first (alternative-specific) portion* that does not vary across observations in the sample and a *second (individual-specific) portion* that varies across observations in the sample. The order condition only applies to the *first portion*. Specifically, the maximum number of covariance terms that can be estimated from the first portion of Ω is given by $S = \frac{J(J-1)}{2} - 1$, where J is the number of choice alternatives. In the current context, the entire variance-covariance matrix does not vary across observations. Therefore, with $J = 3$, S becomes 2.

According to the rank condition, the maximum number of estimable parameters (M_{Rank}) is:

$$M_{Rank} = rank[jacobian[vecu(\Omega_\Delta)]] - 1 \quad (11)$$

where, $vecu(\Omega_\Delta)$ is the function to vectorize the unique elements of Ω_Δ into a column vector. The resulting $vecu(\Omega_\Delta)$ and its Jacobian matrix $jacobian[vecu(\Omega_\Delta)]$ for the error difference variance-covariance matrix in Eq. (10) are:

$$vecu(\Omega_\Delta) = \begin{bmatrix} \beta_{a1}^2\sigma_{\eta_1}^2 + \beta_{a2}^2\sigma_{\eta_2}^2 + \beta_{a3}^2\sigma_{\eta_3}^2 + 2\sigma_\xi^2 \\ \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 + \sigma_\xi^2 \\ \beta_{d1}^2\sigma_{\eta_1}^2 + \beta_{d2}^2\sigma_{\eta_2}^2 + \beta_{d3}^2\sigma_{\eta_3}^2 + 2\sigma_\xi^2 \end{bmatrix} \tag{12}$$

$$\text{and } Jacobian[vecu(\Omega_\Delta)] = \begin{bmatrix} \beta_{a1}^2 & \beta_{a2}^2 & \beta_{a3}^2 & 2 \\ \beta_{a1}\beta_{d1} & \beta_{a2}\beta_{d2} & \beta_{a3}\beta_{d3} & 1 \\ \beta_{d1}^2 & \beta_{d2}^2 & \beta_{d3}^2 & 2 \end{bmatrix}. \tag{13}$$

The rank of this Jacobian matrix is 3. It can be verified that even if there were more than three traffic environment variables with additive errors, the rank would be equal to 3. Therefore, only two parameters can be estimated in the variance-covariance matrix Ω. This suggests that one cannot estimate unique scale parameters associated with the additive perception error terms separately for each of the three traffic environment variables.

The above discussion was for a specific case of 3 choice alternatives. In a general case with J choice alternatives, it can be shown that an additive error specification in choice environment variables results in order and rank conditions that allow the identification of up to $J-1$ parameters in the variance-covariance matrix (Ω) of error terms. Therefore, in contexts with small choice sets (such as the current empirical context with only 3 alternatives), it is not possible to explore additive stochasticity in several choice environment variables that enter the utility functions without interacting with any alternative attributes.

3.3.2. Identification for Multiplicative Specification of Errors in Choice Environment Variables

Employing the multiplicative error specification of Eq. (5) for choice environment variables, the utility structure in Eq. (6) may be written as:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}\tau_{q1}) + \beta_{d2}(x_{q2}\tau_{q2}) + \beta_{d3}(x_{q3}\tau_{q3}) + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{14}$$

Let the random components of the above utility terms be written as:

$$\begin{aligned}\varsigma_{qa} &= \beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa}\\ \varsigma_{qd} &= \beta_{d1}(x_{q1}\tau_{q1}) + \beta_{d2}(x_{q2}\tau_{q2}) + \beta_{d3}(x_{q3}\tau_{q3}) + \xi_{qd}\\ \varsigma_{qs} &= \xi_{qs}\end{aligned} \tag{15}$$

As discussed earlier, the random error τ_{qk} should be specified to have an expected value of one, i.e., $E[\tau_{qk}]=1$. Further, the sign of a perceived variable value can be assumed to be the same as that of the observed value. And physical quantities such as distances and time ought to be positive. Therefore, distributions with support on the positive side of the real line are suitable for τ_{qk}.

To continue the discourse, let us assume that: (a) the perception error term τ_{qk} is lognormally distributed with location parameter $\mu_{\tau k}$, scale parameter σ_{τ_k}, and mean 1, and (b) the kernel error terms ξ_{qj} are IID Gumbel with location parameter zero and scale parameter g_ξ (variance $\sigma_\xi^2 = \pi^2 / 6g_\xi^2$). For the expected value of the lognormally distributed τ_{qk} term to be 1 its parameters should fulfil the restriction that $\mu_{\tau_k} = \frac{-\sigma_{\tau_k}}{2}\ \forall k = 1,2,3$. With these assumptions, one can derive the variance-covariance matrix of the stochastic utility terms as below:

$$\Omega = \begin{bmatrix} Var(\varsigma_{qa}) & Cov(\varsigma_{qa},\varsigma_{qd}) & Cov(\varsigma_{qa},\varsigma_{qs})\\ Cov(\varsigma_{qa},\varsigma_{qd}) & Var(\varsigma_{qd}) & Cov(\varsigma_{qd},\varsigma_{qs})\\ Cov(\varsigma_{qa},\varsigma_{qs}) & Cov(\varsigma_{qd},\varsigma_{qs}) & Var(\varsigma_{qs})\end{bmatrix} \tag{16}$$

where,

$$\begin{aligned}&Var(\varsigma_{qa}) = (\beta_{a1}x_{q1})^2[\exp(\sigma_{\tau1}^2)-1] + (\beta_{a2}x_{q2})^2[\exp(\sigma_{\tau2}^2)-1] + (\beta_{a3}x_{q3})^2[\exp(\sigma_{\tau3}^2)-1] + \sigma_\xi^2\\ &Var(\varsigma_{qd}) = (\beta_{d1}x_{q1})^2[\exp(\sigma_{\tau1}^2)-1] + (\beta_{d2}x_{q2})^2[\exp(\sigma_{\tau2}^2)-1] + (\beta_{d3}x_{q3})^2[\exp(\sigma_{\tau3}^2)-1] + \sigma_\xi^2\\ &Var(\varsigma_{qs}) = \sigma_\xi^2\\ &Cov(\varsigma_{qa},\varsigma_{qd}) = \beta_{a1}\beta_{d1}\left(x_{q1}\right)^2\exp(\sigma_{\tau1}^2) + \beta_{a2}\beta_{d2}\left(x_{q2}\right)^2\exp(\sigma_{\tau2}^2) + \beta_{a3}\beta_{d3}\left(x_{q3}\right)^2\exp(\sigma_{\tau3}^2)\\ &Cov(\varsigma_{qa},\varsigma_{qs}) = 0\\ &Cov(\varsigma_{qd},\varsigma_{qs}) = 0.\end{aligned}$$

The corresponding covariance matrix of error differences with respect to the base alternative is:

$$\Omega_\Delta = \begin{bmatrix} Var(\varsigma_{qa}) + Var(\varsigma_{qs}) - 2Cov(\varsigma_{qa},\varsigma_{qs}) & Cov(\varsigma_{qa},\varsigma_{qd}) + Var(\varsigma_{qs})\\ Cov(\varsigma_{qa},\varsigma_{qd}) + Var(\varsigma_{qs}) & Var(\varsigma_{qd}) + Var(\varsigma_{qs}) - 2Cov(\varsigma_{qd},\varsigma_{qs})\end{bmatrix} \tag{17}$$

Observe from the above variance-covariance matrix that, unlike in the case of additive error specification, the measurements x_{qk} enter the variance-covariance matrix and render its elements to vary across observations. Such additional information derived from the variation of the covariance matrix across observations helps in uncovering stochasticity (σ_{τ_k} parameters) for as many traffic environment variables as needed, just as the typical mixed logit model allows the estimation of random coefficients on any number of alternative attributes (Walker, 2001). In sum, the multiplicative error specification, in theory, allows the estimation of stochasticity in any number of choice environment variables entering the utility functions – as long as the variables have a statistically significant influence on the choice outcome. Of course, empirical identifiability issues might arise if one attempts to uncover stochasticity in too many variables.

A final note is in order here regarding the suitability of our findings on the econometric identifiability of multiplicative vs. additive stochastic specification for errors in variables that do not vary across alternatives. As mentioned earlier, the findings in this section on parameter identifiability are specific to contexts when choice environment variables enter the utility functions without interacting with alternative attributes. On the other hand, in situations when the choice environment variables (also, demographic variables) are interacted with alternative attributes that vary across alternatives and individuals (e.g., income interacted with mode-specific costs in mode choice models), such interactions render the resulting variance-covariance matrix of the random utility terms to vary across individuals. This can potentially help improve the identifiability of the parameters of additive stochastic specification on choice environment variables that are interacted with alternative attributes – if the coefficients on such interactions are specified as deterministic. Of course, this note is based purely on statistical considerations of parameter identifiability. However, for variables representing physical quantities (such as space gaps) that involve perception errors of the decision-makers, psychophysical theories of human perception favour the multiplicative stochastic specification – regardless of how the variables enter the utility functions.

3.4. Comparison with the Error Components Specification

It is important to note that the above-discussed multiplicative specification allows separate identification of stochasticity for each choice environment variable (i.e., one can estimate σ_{τ_k} separately for each x^*_{qk}) that has a statistically significant influence on the choice outcome.

Therefore, unlike in Díaz et al. (2015), there is no need to combine the stochasticity of all variables into alternative-specific error components. This helps in (1) the interpretation of the uncovered stochasticity separately for each choice environment variable and (2) comparing variability due to perception errors in different choice environment variables.

Note that stochasticity in choice environment variables introduces differential variance across choice alternatives (because of alternative-specific coefficients on these variables) and correlations among utility functions (because of common stochastic variables entering different utility functions). Therefore, one might suggest that error components that allow heteroscedasticity across alternatives or correlation among choice alternatives can help capture stochasticity due to perception errors in choice environment variables. This is unlikely because multiplicative stochasticity is not easily separable from the deterministic utility function into error components. Therefore, existing variants of mixed logit models, such as error component models, may not be suitable to accommodate such stochasticity. This is demonstrated through both simulated data and empirical data in Sections 5 and 6, respectively.

3.5. Comparison with the Random Coefficients Specification

In the context of the multiplicative stochasticity specification as in Eq. (5) for choice environment variables, the stochasticity in the τ_{qk} term might be confounded with random heterogeneity in β_{ik} (i.e., drivers' sensitivity to the variable x^*_{qk}), even if the intent of including it is to capture stochasticity in x^*_{qk}. One can see this by substituting $x_{qk}\tau_{qk}$ for x^*_{qk} in the utility functions, as below:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \sum_{k=1}^{K} \beta_{ak} x_{qk} \tau_{qk} + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \sum_{k=1}^{K} \beta_{dk} x_{qk} \tau_{qk} + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{18}$$

There are two possible cases in the context of random coefficients in the above utility functions. The first case is when β_{ak} and β_{dk} are random but uncorrelated. The risk of confounding between the stochasticity in τ_{qk} and that in β_{ak} and β_{dk} is low in this case. This is because β_{ak} and β_{dk} are alternative-specific, and their distributions would be different, even if their standard deviations are of the same value. On the other hand, the distribution of x^*_{qk} is the same regardless of which alternative's utility function it enters because x_{qk}^{*} does not vary across alternatives. Therefore, if there are no strong reasons to believe that β_{ak} and β_{dk} are correlated, one can safely interpret τ_{qk} as representing stochasticity due to perception errors than random heterogeneity in β_{ak} or β_{dk}.

The second, more general case is when β_{ak} and β_{dk} are correlated random coefficients (CRC). There is a high risk of confounding in this case because the correlation between β_{ak} and β_{dk} can pick up the stochasticity in τ_{qk}, which reduces the need for (and identifiability of) a separate τ_{qk} term. Therefore, in situations with both stochastic variables (x_{qk}^{*}) and correlated random coefficients on those variables, the correlated random coefficients (CRC) model structure without additional τ_{qk} terms might suffice. Alternatively, an uncorrelated random coefficients model with additional τ_{qk} terms may be explored as well. In either case, one cannot separately identify the correlations between random coefficients from stochasticity in the variables. Nonetheless, either of these models would work better than a model with only multiplicative stochastic terms (τ_{qk}) and no random coefficients.[4]

However, an important question in this context is whether the CRC model can be used if the primary source of stochasticity is in x_{qk}^{*}, not in its coefficients. In such situations, although the CRC model is a more general structure that subsumes the multiplicative stochastic variable model as a special case, the former model would run into parameter (un)identifiability problems during estimation. To understand this better, consider the following utility functions with correlated random coefficients that are lognormally distributed:

[4] For the same reason, the CRC model without the τ_{qk} term can be used to represent the first case with uncorrelated random coefficients (β_{ak} and β_{dk}) and stochasticity (τ_{qk}) in the choice environment variable, assuming the distributional assumptions allows recasting of one model to the other. Such a model would have the same number of parameters (means and standard deviations of β_{ak} and β_{dk}, and a correlation parameter) as a model that separately estimates uncorrelated random coefficients and stochasticity in x_{qk}^{*}, *ceteris paribus*. In this case, the CRC model structure would not be superior to a model with uncorrelated random coefficients and stochastic variables. If the analyst believes the correlation among random coefficients in a CRC model is due to stochasticity in the corresponding variable, then the latter model should be used for interpretation.

$$
\begin{aligned}
U_{qa} &= \beta_{a0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{akRC} - \sigma_{ak}\Phi^{-1}\left[1-\Phi\left(Z_{qak}\right)\right]\right) x_{qk} + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{dkRC} - \sigma_{dk}\Phi^{-1}\left[1-\Phi\left(\rho_k Z_{qak} + \left(\sqrt{1-\rho_k^2}\right) Z_{qdk}\right)\right]\right) x_{qk} + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned}
\tag{19}
$$

In the above utility structure, the correlations are between the random coefficients of x_{qk} in the utility functions of alternatives a and d (Z_{qak} and Z_{qdk} are standard normal variates). The lognormally distributed random coefficients are: $\exp\left(\beta_{akRC} - \sigma_{ak}\Phi^{-1}\left[1-\Phi\left(Z_{qak}\right)\right]\right)$ and $\exp\left(\beta_{dkRC} - \sigma_{dk}\Phi^{-1}\left[1-\Phi\left(\rho_k Z_{qak} + \left(\sqrt{1-\rho_k^2}\right) Z_{qdk}\right)\right]\right)$, respectively. In this utility structure, when the following restrictions are imposed: $\sigma_{ak} = \sigma_{dk} = \sigma_k$ and $\rho_k = 1$, it implies that the random coefficients on x_{qk} in the utility functions of alternatives a and d are exactly the same (i.e., perfectly correlated). In such a special case, when the expected values of the random components of the coefficients become 1, i.e. $E\left[\exp\left(-\sigma_{ak}\Phi^{-1}\left[1-\Phi\left(Z_{qak}\right)\right]\right)\right] = 1$ and $E\left[\exp\left(-\sigma_{dk}\Phi^{-1}\left[1-\Phi\left(\rho_k Z_{qak} + \left(\sqrt{1-\rho_k^2}\right) Z_{qdk}\right)\right]\right)\right] = 1$, the utility structure simplifies as below:

$$
\begin{aligned}
U_{qa} &= \beta_{a0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{akRC}\right) \times \exp\left(\sigma_k Z_{qak} - 0.5\sigma_k^2\right) \times x_{qk} + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{dkRC}\right) \times \exp\left(\sigma_k Z_{qak} - 0.5\sigma_k^2\right) \times x_{qk} + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned}
\tag{20}
$$

Assuming $\exp\left(\sigma_{ak} Z_{qak} - 0.5\sigma_k^2\right) = \tau_{qk}$, $\exp\left(\beta_{akRC}\right) = \beta_{ak}$, and $\exp\left(\beta_{dkRC}\right) = \beta_{dk}$, the above utility structure simplifies to that in Eq. (18), where τ_{qk} is viewed as a multiplicative error on x_{qk}.

In sum, the proposed model with multiplicative stochasticity (EIV) on a choice environment variable is a special case of a CRC model with perfectly correlated random coefficients on that variable. Given this result, a natural question is what is the need for the proposed multiplicative EIV model when it is a special case of a more general CRC model? To answer this question, it is important to note that one cannot estimate a CRC model when the primary source of stochasticity is multiplicative EIV and not random parameters (for variables that are not alternative-specific). The estimation would lead to identification problems because the single source of stochasticity (multiplicative EIV) is not sufficient to identify different random parameters that are perfectly correlated, have the same scale parameters, and have an expected value of 1 (this is demonstrated using simulated data in Section 5.4). In such situations, a CRC model would be identified only if the above-mentioned constraints are imposed. But such a model is the same as the model with multiplicative EIV and no random coefficients. Therefore, when the data has only multiplicative stochasticity in attributes and no random heterogeneity in response to those attributes, the multiplicative EIV specification should be preferred.[5]

3.6. Alternative Distributions for Multiplicative Errors in Choice Environment Variables

As indicated earlier, choice environment variables representing physical quantities such as distance, time, and speed cannot be negative. Also, it is reasonable to assume that people do not perceive positive relative speeds as negative or vice versa. Therefore, the distributions used for multiplicative errors in such choice environment variables should not flip the sign of the observed value. Further, the expected value of the distribution ought to be normalized to 1 for identification and for zero bias in perception. The statistical literature has a variety of distributions with support on the positive semi-infinite interval. In this study, we explored the following three distributions: (1) the power lognormal (PLN) distribution, which subsumes the lognormal distribution as a special case, (2) the Weibull distribution, which subsumes the Rayleigh distribution and the exponential distribution as special cases, and (3) The Fréchet distribution.

Table B.1 in Appendix B provides a brief overview of each of these distributions, including their density function, permissible ranges of parameter values and support of the distribution. In addition, the expression for the location parameter (μ) is provided as a function of the scale parameter (σ) and other (if any) parameters of the distribution – to normalize the expected value

[5] In general, mixed multinomial outcome models with correlated random coefficients (CRC) may be viewed as more general versions of models with a single source of unobserved heterogeneity common to multiple parameters or variables. In such situations, although the CRC model subsumes the specifications with a single source of unobserved heterogeneity as special cases, attempts to estimate the CRC model will likely run into identification problems as it involves more parameters than those necessary to identify a single source of heterogeneity. One such situation is the model discussed in this paper, where the primary source of unobserved heterogeneity arises due to human errors in the perception of physical quantities (more generally, errors in variables that do not vary across alternatives). Another such situation arises when random heterogeneity in the scale of utility functions across individuals is the primary source of unobserved heterogeneity (see Hess and Train (2017) for a discussion on the relationship between scale heterogeneity and correlation in random coefficients). Even in the latter situation, estimating a CRC model would be encountered with parameter identification issues.

of the distribution to 1. The expressions for inverse CDF function and standard deviation are provided when the expected value of the distribution is equal to 1. The inverse CDF function is useful for simulating the corresponding distributions in MSL estimation. The standard deviation is useful for comparing variations in perception errors of different variables.

The first application of PLN distribution in the choice modelling literature was by Bhat and Lavieri (2018), who used it for random coefficients on travel time and travel cost variables. Other than the location parameter and scale parameter, a power parameter (p) governs the thickness of the distribution's tail. At $p = 1$, the distribution becomes lognormal. As the value of p increases beyond 1, the tail of the distribution becomes thinner. This property, as discussed in Bhat and Lavieri (2018), makes it easier to estimate the parameters of a PLN distribution (when $p > 1$) compared to those of a lognormal distribution. Note also that the location parameter can be any real value for PLN and lognormal distributions while keeping the support to be strictly positive. Thanks to this property, there is no need to constrain the value of σ (i.e., the analyst can let the data decide its value).

For the other distributions reviewed in the table, however, the location parameter (μ) cannot be negative. This, combined with the normalization that the expected value is 1, imposes a constraint on the permissible values of the scale parameter. Furthermore, for these distributions μ is the minimum value that a random variable can take. All these constraints make it difficult to estimate models with such distributions for multiplicative errors in choice environment variables. This is because estimating a scale parameter (while $\mu \geq 0$) implies that the distribution of the perception error does not allow values less than μ. This implies that people do not underestimate choice environment variables below what is permissible by μ – an assumption that cannot be easily justified. Instead, setting μ a prespecified value fixes (restricts) the scale parameter because of the normalization that the expected value is 1. Therefore, the PLN distribution is likely to be more suitable than the other distributions for multiplicative stochasticity.

4. DATA

The main source of data used in this study – both for simulation experiments and empirical analysis – comes from a 30-minute video of a heterogeneous traffic stream on an urban arterial stretch of 245 m in the city of Chennai, India. Kanagaraj et al. (2015) processed the raw video data into vehicle trajectories and made it available for use by the research community. Their vehicle trajectory data includes information on the type and dimensions of each vehicle in the video and the space-time trajectory of each vehicle at a 0.5 s resolution, including the position, speed, and acceleration/deceleration values in both the longitudinal and lateral dimensions (to the roadway).

In another study, Nirmale et al. (2021) further processed this data to identify a rectangular influence zone around each vehicle at each time step of 0.5 s, as shown in Figure 1. The influence zone was of length 30 m (plus the vehicle's length), with the road boundaries defining the width of the influence zone. In this figure, the vehicle in red colour and marked SV is the subject vehicle. The influence zone around the SV is divided into five compartments. The space directly ahead of SV is labelled the middle front (MF) compartment, the space ahead to the left of SV is labelled the left front (LF) compartment (similarly, the space ahead to the right of SV is called the RF compartment), and the adjacent space to the left of SV is called the left side (LS) compartment (space to the right side of SV is called the RS compartment). Vehicles in each of these compartments are labelled accordingly as shown in the legend of the figure.

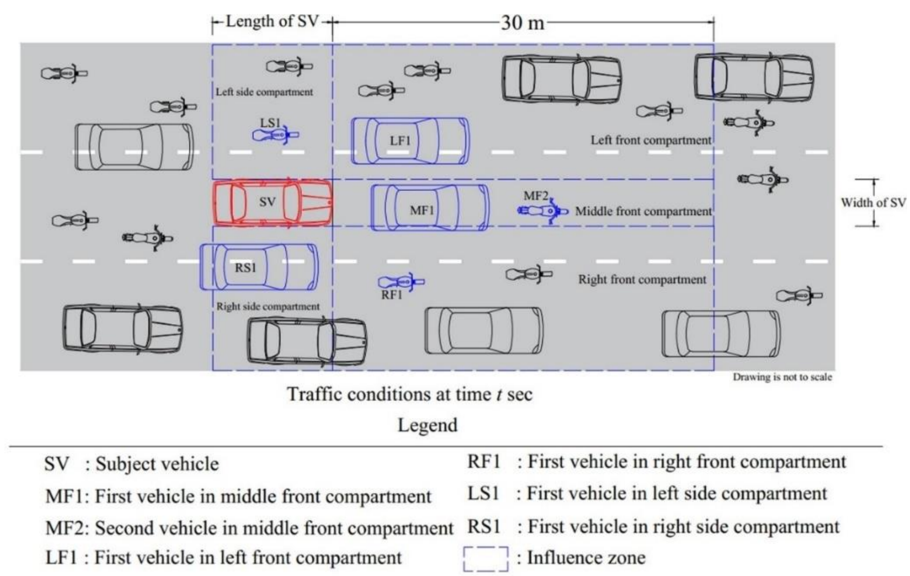

Figure 1 Structure of influence zone around a subject vehicle (Source: Nirmale et al., 2021)

At each 0.5 s time instance t for each subject vehicle, the following data were identified: (a) the longitudinal and lateral position, speed, and acceleration/deceleration/steady speed states of the subject vehicle at the time instance t and at $t-0.5$ s (note: 0.5 s is considered the reaction time, based on an analysis by Nirmale et al., 2021) (b) all other vehicles and their characteristics (type and dimensions) and infrastructure elements within the influence zone at $t-0.5$ s, and (c) traffic environment variables such as space gaps and relative speeds of the SV with respect to other vehicles and infrastructure elements in the influence zone at $t-0.5$ s. The final data comprises 17,852 observations from 749 passenger cars. Of these records, a subset was chosen for simulation experiments and empirical analysis. The remaining data were set aside for validation purposes.

5. SIMULATION STUDY

We carried out simulation experiments for the following purposes: (a) to evaluate the ability to identify and retrieve parameters of the proposed multiplicative EIV model using MSL estimation, (b) to compare the performance of the proposed multiplicative EIV model against the typically used mixed logit models with random coefficients or error components when the data generation process (DGP) has stochasticity in explanatory variables (x_{qk}^*) but not in their coefficients (β_{ik}), (c) to evaluate alternative model structures when the DGP has stochasticity in the coefficients of choice environment variables (β_{ik}), but not in the variables themselves (x_{qk}^*), and (d) to develop guidelines for which model structure to use when. This section describes the simulation setup, presents the results, and discusses findings from the simulation experiments.

5.1. Experimental Design for Synthetic Dataset Generation

To generate synthetic datasets for the simulation experiments, we used a subset of 8,540 observations from the earlier-described empirical data for measurements of the explanatory variables (x_{qk}). The data were used to estimate simple empirical models for the proposed model structure with multiplicative perception errors for the traffic environment variables. Next, the parameter estimates of these empirical models were assumed as '*true*' parameter values and applied back on the same empirical data to calculate the utility function values for each choice alternative – acceleration, deceleration, and maintain same speed. To do so, the random components of the utility functions were simulated according to their assumed distributions. Subsequently, the alternative with the highest utility value was denoted as the chosen alternative.

The following four variables were assumed to enter the utility function of a subject vehicle (SV) q's driver: (a) speed of the SV (x_{q1}^*), (b) perceived longitudinal space gap between SV and MF1 (x_{q2}^*), (c) perceived relative speed between SV and MF1 (x_{q3}^*), and (d) perceived relative speed between SV and LF1 (x_{q4}^*). Among these four variables, it was assumed that the SV's driver would know her/his vehicle's speed accurately (i.e., $x_{q1}^{*} = x_{q1}$). The other three variables were considered stochastic due to multiplicative perception errors. That is, $x_{qk}^{*} = x_{qk}\tau_{qk}\ (k = 2,3,4)$, where x_{qk} is the observed value of the k^{th} traffic environment variable and τ_{qk} is the multiplicative error term assumed to be power lognormal (PLN) distributed. The resulting utility functions for acceleration, deceleration, and maintain same speed decisions – U_{qa}, U_{qd}, U_{qs} – are as below:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \beta_{a4}(x_{q4}\tau_{q4}) + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \beta_{d2}(x_{q2}\tau_{q2}) + \beta_{d3}(x_{q3}\tau_{q3}) + \beta_{d4}(x_{q4}\tau_{q4}) + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{21}$$

In the above utility functions, β_{ak} and β_{dk} $(k = 1,2,3,4)$ are coefficients of the traffic environment variables ($x_{qk}^{*};\ k = 1,2,3,4$) in the acceleration and deceleration utility functions, respectively. The parameters of the PLN distributed terms $(\tau_{qk};\ k = 2,3,4)$ for perception errors are the scale parameter σ_k and power parameter p_k, with the location parameter μ_k set to be equal to $-\ln\left(\int_0^1 \exp(-\sigma_k \Phi^{-1}(y^{1/p_k}))dy\right)$ to ensure unit expected value for the distribution. Finally, ξ_{qa}, ξ_{qd} and ξ_{qs} are IID standard Gumbel error terms. Table 1 (in its second column) provides the true values of β_{ik} and σ_k used for generating the synthetic datasets. The power parameter p_k was set to be 3 for all three stochastic variables $x_{qk}^{*}\ (k = 2,3,4)$. For brevity, the resulting model is labelled the ML-ME-PLN model to indicate that the multiplicative errors are specified to be PLN distributed.

A total of 115 datasets of 8,540 records each were generated for the ML-ME-PLN model. The average of the sample shares of acceleration, deceleration, and maintain same speed choices simulated across these datasets are 41.9%, 45.7%, and 12.4%, respectively, which are similar to those observed in the empirical data.

5.2. Parameter Recovery of the Proposed ML-ME-PLN Model

The following performance metrics were computed to evaluate the accuracy and precision with which the parameters of the proposed model were recovered using the MSL estimation method:

- Absolute percentage bias (APB): Estimate parameters for each of the 115 datasets and compute the mean of the estimates across all datasets. For each parameter, $APB = \left|\frac{\text{mean estimate - true value}}{\text{true value}}\right| \times 100$.
- Finite sample standard error (FSSE): FSSE, a measure of the empirical standard error, is the standard deviation of the parameter estimates across the 115 datasets.
- Asymptotic Standard Error (ASE): ASE is the mean of standard error across all datasets.
- Root mean squared error (RMSE) = $\sqrt{\left(\text{mean estimate - true value}\right)^2 + FSSE^2}$

A summary of the above performance metrics is presented in Table 1 for each parameter. As can be observed from the table, the proposed model was able to recover parameters accurately even when only 200 Halton draws were used to simulate the distributions of perception errors. The mean APB value across all parameters is 4.75%, which is small. The low FSSE values suggest a high empirical (finite-sample) efficiency in recovering the parameters. While the FSSE values for the scale parameters of perception error distributions are relatively higher than those for the coefficients of explanatory variables, their absolute values are small. Also, the ASE values are close to the corresponding FSSE values, except for the scale parameter σ_4, suggesting that the ASE values provide a good approximation to the FSSE values in finite samples. A high ASE value for σ_4 (relative to its FSSE value) may be because of the use of empirical data from the field for measurements of the explanatory variables.[6] Nevertheless, the RMSE measure, which combines the bias and efficiency measures into a single metric across all parameters, is small, suggesting very good parameter recovery. Importantly, these results demonstrate that it is possible to separately identify stochasticity in each choice environment variable through the proposed specification rather than combining the stochasticity of all variables into a few error components. This helps in obtaining insights into which variables are associated with greater variability than others.

[6] When we conducted additional simulations using fully simulated data (i.e., the measurements of explanatory variables, too, were simulated), we observed accurate and efficient parameter recovery for all parameters and did not encounter issues such as the ASE and FSSE values being quite different.

Table 1 Metrics of Parameter Recovery for the ML-ME-PLN Model

Parameters	**True value**	**Mean**	**APB (%)**	**FSSE**	**ASE**	**RMSE**
β_{a0}	2.010	1.978	1.585	0.230	0.265	0.232
β_{d0}	-1.320	-1.375	4.172	0.233	0.298	0.240
β_{a1}	-0.100	-0.094	5.549	0.024	0.027	0.025
β_{d1}	0.230	0.240	4.147	0.024	0.029	0.026
β_{a2}	0.030	0.026	14.235	0.012	0.011	0.013
β_{d2}	-0.120	-0.120	0.315	0.045	0.050	0.045
β_{a3}	0.330	0.334	1.346	0.094	0.095	0.094
β_{d3}	-0.390	-0.417	6.828	0.103	0.118	0.107
β_{a4}	0.100	0.094	6.228	0.026	0.026	0.027
β_{d4}	-0.020	-0.021	5.804	0.015	0.018	0.015
σ_2	2.760	2.535	8.152	0.339	0.365	0.407
σ_3	2.030	2.038	0.389	0.325	0.354	0.325
σ_4	1.250	1.212	3.028	0.416	1.136	0.418
Mean value	--	--	4.752	0.145	0.215	0.152

5.3. Performance of Alternative Mixed Logit Models with Random Coefficients or Error Components when the Primary Source of Stochasticity in DGP is in x^*_{qk}, Not in β_{ik}

In addition to the proposed ML-ME-PLN model with the utility specification as in Equation (21), the following alternative ML models were estimated on the same simulated data from Section 5.1:

(a) ML model with PLN distributed uncorrelated random coefficients (labelled ML-RC-PLN), with the following utility structure:

$$
\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \sum_{k=2,3,4} (\tau_{qak}) x_{qk} + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \sum_{k=2,3,4} (\tau_{qdk}) x_{qk} + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned}
\tag{22}
$$

Here, τ_{qak} and τ_{qdk} are PLN distributed (and uncorrelated) random coefficients on x_{qk} in the acceleration and deceleration utility functions, respectively. The location and scale parameters of these random coefficients are to be estimated.

(b) ML model with PLN distributed and correlated random coefficients (labelled ML-CRC-PLN). In this model, the utility equations would look similar to those in Eq. (22), except that the PLN distributed random coefficients τ_{qak} and τ_{qdk} are correlated with a correlation parameter ρ_k. The correlated PLN distributed terms can be expressed as:

$$\tau_{qak} = \exp\left(\beta_{ak} - \sigma_{k1}\Phi^{-1}\left[\left(1-\Phi\left(Z_{qk1}\right)\right)^{1/p_{k1}} \right] \right)$$
$$\tau_{qdk} = \exp\left(\beta_{dk} - \sigma_{k2}\Phi^{-1}\left[\left(1-\Phi\left(\rho_k Z_{qk1} + \left(\sqrt{1-\left(\rho_k\right)^2}\right) Z_{qk2} \right)\right)^{1/p_{k2}} \right] \right) \tag{23}$$

Recall (from Section 3.5) that the correlated random coefficients model subsumes the multiplicative EIV model as a special case when the corresponding random coefficients are perfectly correlated (with the same scale parameters) and have an expected value of 1.

(c) ML model with error components for correlation between the utility functions of acceleration and deceleration alternatives, but no random coefficients or stochastic variables (labelled ML-EC-rho), as below:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \sum_{k=2,3,4} \beta_{ak} x_{qk} + \tau_{qad} + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \sum_{k=2,3,4} \beta_{dk} x_{qk} + \tau_{qad} + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{24}$$

Here, τ_{qad} is a normal distributed error component with mean zero (and scale to be estimated) to allow correlation between the acceleration and deceleration utility functions.

(d) ML model with error components for heteroscedasticity across choice alternatives but no random coefficients or stochastic variables (labelled ML-EC-het), as below:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \sum_{k=2,3,4} \beta_{ak} x_{qk} + \tau_{qa} + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \sum_{k=2,3,4} \beta_{dk} x_{qk} + \tau_{qd} + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{25}$$

Here, τ_{qa} and τ_{qd} are normal distributed error components with mean zero (and scale parameters to be estimated) to allow heteroscedasticity across utility functions.

To compare the proposed ML-ME-PLN model vis-à-vis alternative ML models with random coefficients or error components, we compared the model fit using the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC). Table 2 presents the percentage of simulated datasets (of the 115 datasets) for which each alternative model structure showed better AIC or BIC values than others. As can be observed from this table, the proposed ML-ME-PLN model provided a better fit than all other ML models in more than 92% of the datasets. The ML model with PLN distributed but uncorrelated random coefficients was better for less than 7% of the datasets. Interestingly, the model with PLN distributed and correlated random coefficients never performed better than the ML-ME-PLN model (more on this soon). And neither of the error components models performed better in any of the 115 datasets. The above results suggest that typically used ML models with random coefficients on a choice environment variable or those with error components do not help capture stochasticity in that variable (if multiplicative EIV, not random coefficients, is the predominant source of stochasticity). In addition, such models lead to inferior fit to data and potentially biased parameter estimates.

Table 2 Performance of alternative mixed logit models when the data generation process has stochasticity in choice environment variables

Preferable model over the other models	No. of datasets according to AIC (%)	No. of datasets according to BIC (%)
ML-ME-PLN model (PLN distributed stochastic variables)	94.78	93.04
ML-RC-PLN model (PLN distributed uncorrelated random coefficients)	5.22	6.96
ML-CRC-PLN model (PLN distributed correlated random coefficients)*	0.00	0.00
ML-EC-rho model for correlation between U_{qa} and U_{qd}	0.00	0.00
ML-EC-het model for heteroscedasticity across alternatives	0.00	0.00
Total number of simulated datasets	115	

* Parameter identification problems were faced when estimating the ML-CRC-PLN model

Importantly, the ML-CRC-PLN model, even though it is a more general model that subsumes the DGP (with only multiplicative stochastic variables) as a special case, could not be estimated in most of the 115 datasets. Attempts to estimate this model resulted in non-invertible Hessians or very high standard errors for estimates related to the correlated random coefficients. These manifestations are characteristic of an unidentified model. Further, the correlation parameter (ρ_k) estimates, if the corresponding standard errors could be determined, were of high magnitude (higher than 0.9), indicating near perfect correlation between the corresponding random coefficients across different choice alternatives. These results corroborate our claim in Section 3.5 that the correlated random coefficients model cannot be used when the primary source of stochasticity is predominantly due to multiplicative stochasticity in choice environment variables. In such a situation, the analyst should estimate a simpler model that directly specifies multiplicative stochasticity in choice environment variables than a CRC model.

5.4. Performance of Alternative Mixed Logit Models when Underlying Data has Only Random Coefficients on x_{qk}^* but No Stochasticity in x_{qk}^*

Now we examine which model performs better when the underlying data has random coefficients in the choice environment variables but no stochasticity in those variables. To do this, we simulated 100 sets of datasets – each set includes six datasets simulated assuming six different DGPs as described below (each dataset is of sample size 3,000):

1. DGP1: Two uncorrelated random coefficients on a choice environment variable (in two different utility functions), with scale parameter values close to each other (scale parameters are 1.00 and 1.20, ρ = 0.0).
2. DGP2: Two uncorrelated random coefficients on a choice environment variable, and the scale parameters are not close to each other (scale parameters are 0.80 and 1.50, ρ = 0.0).
3. DGP3: Two correlated random coefficients on a choice environment variable, with scale parameter values close to each other, and the correlation level is high (scale parameters are 1.00 and 1.20, ρ = 0.7).
4. DGP4: Two correlated random coefficients on a choice environment variable, with scale parameter values close to each other, and the correlation level is low (scale parameters are 1.00 and 1.20, ρ = 0.3).
5. DGP5: Two correlated random coefficients on a choice environment variable, their scale parameter values are not close to each other, and the correlation level is high (scale parameters are 0.80 and 1.50, ρ = 0.7).
6. DGP6: Two correlated random coefficients on a choice environment variable, their scale parameter values are not close to each other, and the correlation level is low (scale parameters are 0.80 and 1.50, ρ = 0.3).

Table 3 presents the performance of alternative mixed logit models for all six cases according to the AIC metric. As can be observed, in each of the six cases, for most of the 100 simulated datasets, the true DGP model performs better than the ML-ME-PLN model that specifies multiplicative stochasticity on the variable. These results suggest that when the underlying DGP has random coefficients with or without correlations, a model that specifies only multiplicative stochasticity on the corresponding variables is less likely to pick up such stochasticity. Only for DGP3, where the random coefficients have similar standard deviation values and high correlation, the multiplicative stochasticity model showed better performance in 29% of the datasets. That is, the multiplicative error model is likely to pick up correlated random sensitivities to an attribute only if the correlation is high and the standard deviations of random coefficients are similar.[7]

Table 3 Statistical performance of alternative ML models (according to AIC)

Data generating process	**No. of datasets where ML-CRC-PLN model is preferred over the other models**	**No. of datasets where ML-RC-PLN model is preferred over the other models**	**No. of datasets where ML-ME-PLN model is preferred over the other models**
DGP1	--	87	13
DGP2	--	95	5
DGP3	71	--	29
DGP4	82	--	18
DGP5	81	--	19
DGP6	90	--	10

5.5. Guidance for Model Selection

Based on the conceptual discussions in Section 3 and the simulation experiments in this section, here we provide a few guidelines to help the analyst decide which model structure to work with – for choice environment variables that do not vary across alternatives.

- First, in addition to the basic MNL, if the analyst believes the presence of stochasticity due to random coefficients, or EIV, or both, then estimate all three models – a random coefficients model without correlations (RC model), a CRC model considering correlated random coefficients across different choice alternatives, and a multiplicative EIV model without random coefficients on the variables with errors. One may also estimate a multiplicative EIV model with uncorrelated random coefficients on the variables with errors. However, such a model can be recast as a CRC model if the distributional assumptions allow doing so.

[7] We simulated another set of 100 datasets with two uncorrelated random parameters that have the same standard deviation value, along with stochasticity on the variable with random coefficients. For these datasets, a model with only stochastic variables did not perform as well as a model with uncorrelated random coefficients and stochasticity on the variable. In fact, we were able to recover the model parameters very well for the latter model that reflects the DGP. These results, combined with the other results in this section suggest that the multiplicative error model is unlikely to pick up random heterogeneity in correlated coefficients unless the correlation is high and standard deviations are of similar value.

To be sure of our conclusions in this section, we also compared the statistical fit of alternative models using the BIC metric as well. The BIC metric favoured the ML-ME-PLN model with multiplicative stochastic variables more often than the AIC metric. This is because BIC penalizes complex models (i.e., models with more parameters) more heavily than AIC (Bishop, 2006). Given a family of models, including the true model, the probability that BIC will favour the correct model approaches one as the sample size tends to infinity (Hastie et al., 2009). Since we used a sample size of only 3000 for our simulated datasets, and since it is known that the BIC metric penalizes complex models more heavily than the AIC metric, we used the AIC metric for our evaluation.

- Considering that most empirical research involves moderate-sized datasets of a few thousand samples or less, use AIC to determine a preferred model structure. In addition to data fit metrics such as AIC, use the following guidelines to select a model structure and its interpretation.
- If the CRC model estimation shows signs of unidentifiability (as discussed in Section 5.4) and the correlation parameter estimate is of high value for a choice environment variable under consideration, there is a high likelihood that the EIV for that variable is the predominant source of stochasticity. The goodness-of-fit metrics, such as AIC, would favour the EIV specification.
- If the CRC model (or an EIV model with uncorrelated random coefficients) offers a better statistical fit than the other models, then the underlying DGP may be one of the following: (1) correlated random sensitivities to the variable under consideration, or (2) uncorrelated random sensitivities to the variable in addition to EIV in the variable, or (3) both correlated random sensitivities and EIV. In such a case, the analyst should use their judgement from the empirical context to determine if the correlations are due to correlated sensitivities on a variable, or EIV, or both. For example, in the driver behaviour context, it is unlikely that unobserved sensitivities to variables such a space gaps and relative speeds have a positive correlation between the acceleration and deceleration choice alternatives. Several unobserved factors, such as drivers' aggressiveness, are likely to be associated with opposite preferences between acceleration and deceleration decisions. So, a positive correlation between random coefficients of such choice alternatives, if any, is likely due to drivers' perception errors but not due to correlated unobserved sensitivities. On the other hand, if the variable under consideration is the type of the lead vehicle in driver behaviour models or traveller's age in mode choice models, the likelihood of EIV is small, for vehicle type (age) can be perceived (measured) accurately. Further, given the close relationship between the different model structures discussed here, it is likely in some empirical contexts that the difference in model fit between different models may not be *practically significant*. Therefore, the analyst should combine statistical fit considerations with intuitive considerations based on the knowledge of the empirical context to decide the final model structure and its interpretation.

6. EMPIRICAL ANALYSIS

6.1. Alternative Model Specifications

To incorporate perception errors in traffic environment variables, we estimated only models with the multiplicative specification of perception errors. This is because the additive specification (as discussed in Section 3) is saddled with parameter identifiability problems. A variety of distributions – lognormal, power lognormal (PLN), Rayleigh, Weibull, exponential, and Fréchet – were explored to represent multiplicative perception errors for traffic environment variables. As discussed earlier, the location parameter (μ) of each of these distributions was specified as a function of the scale parameter such that the expected value of the distribution was 1. Doing so made it difficult to estimate models for all distributions except PLN and lognormal distributions for the reasons discussed in Section 3.6. On the other hand, setting μ to zero and imposing an expected value of 1 resulted in an inferior model fit. Such restrictions automatically imply the scale parameter value of the distribution without utilizing empirical data to inform it. Therefore, we explored model specifications with PLN and lognormal distributions for perception errors.

Specifications with lognormal distributions also encountered convergence problems, presumably because of the fat tail of the distribution (Bartels et al., 2006; Bhat and Lavieri, 2018). Therefore, all subsequent empirical analyses of the ML-ME model were conducted with the PLN distribution for multiplicative errors (i.e., the ML-ME-PLN model). To begin with, the power parameter value was fixed at 1.1, and other parameters were estimated. Subsequently, the power parameter was increased in increments of 0.1, and all other parameters were estimated. This was continued to find the maximum of maximum likelihood values among all estimated models.

In addition to the basic MNL models and the proposed ML-ME-PLN model, all alternative ML models discussed in Section 5.3 were estimated. The estimations were carried out on a subset of 9,530 records of the available empirical data. All estimations were carried out using 400 Halton draws to simulate distributions of the stochastic variables (or parameters) other than the IID Gumbel kernel error terms. In addition, all the estimated models were applied to the remaining 8,322 records set aside for validation.

Among all the models estimated, the ML-EC (error component) models did not yield significant error components and were not statistically different from the basic MNL model, corroborating our finding from the simulation experiments that multiplicative perception errors in choice environment variables cannot be captured through the error component models. Among the random coefficients models we estimated, similar to the experience with simulated datasets in Section 5.3, estimation of the correlated random coefficients (ML-CRC-PLN) model showed clear signs of parameter unidentifiability. For example, the parameter estimates of random coefficients on a few traffic environment variables had very high standard errors. Also, different starting values for the parameters resulted in different convergent values with the same log-likelihood value, suggesting a flat likelihood surface. We could estimate correlated random coefficients on only one traffic environment variable – relative speed between SV and the first lead vehicle in the middle front compartment. For this variable, the estimated correlation parameter between the random coefficients in acceleration and deceleration utility functions was not statistically different from 1. The near-perfect correlation suggests stochasticity in the variable (not in its coefficients).

6.2. Goodness of Fit in Estimation and Validation Datasets

Table 4 summarises the performance metrics of the best-fitting specifications of all other models estimated in this study on both estimation and validation datasets. In the estimation dataset, the log-likelihood ratio (LLR) tests to compare each of the ML models against the MNL model suggest that the latter model can be rejected at least at a 95% confidence level. Among all the ML models, the proposed ML-ME-PLN model with multiplicative stochasticity provides the best AIC, and rho-square values in the estimation data. Further, we performed a non-nested hypothesis test proposed by Horowitz **(**1983**)** to compare the proposed ML-ME-PLN model with each of the other ML models. In this test, the null hypothesis that the model with a lower rho-squared value is the true model is rejected at the significance level given by:

$$\text{Significance Level} = \Phi\left[-\left(-2\left(\rho_H^2-\rho_L^2\right)\times LL(0)+\left(K_H-K_L\right)\right)^{\frac{1}{2}}\right] \tag{26}$$

where, ρ_L^2 is the adjusted likelihood ratio index for the model with the lower value, ρ_H^2 is the adjusted likelihood ratio for the model with the higher value, K_H and K_L are the number of parameters in models H and L, respectively, and Φ is the standard normal cumulative distribution function. Using this test, the null hypotheses that the ML-CRC-PLN and ML-RC-PLN are the true models were rejected at a significance level smaller than 0.001. All these results suggest that the ML-ME-PLN model provides the best fit to the empirical data. Findings from the application of all the estimated models to the validation dataset are similar, with the ML-ME-PLN model providing better predictive metrics than other models. These results suggest that allowing for perception errors in traffic environment variables is more important than allowing unobserved heterogeneity in drivers' response to those variables, at least in the current empirical context.

Table 4 Goodness-of-fit measures of various models estimated in this study

Goodness-of-fit measures in estimation data (N=9,530)				
Measures	**MNL model**	**ML-ME-PLN model**	**ML-RC-PLN model**	**ML-CRC-PLN model**
Log-likelihood at zero	-10469.8	-10469.8	-10469.8	-10469.8
Log-likelihood at constants	-9316.9	-9316.93	-9316.9	-9316.9
Log-likelihood at convergence (L)	-8202.2	-8174.3	-8184.1	-8192.3
Number of parameters (K)	21	25	23	23
LLR w.r.t. MNL (degrees of freedom)	--	55.8 (df = 4)	36.3 (df = 2)	19.9 (df = 2)
AIC value $[2K - 2\ln(L)]$	16446.4	16398.5	16414.1	16430.5
Adj rho-square w.r.t. constants model	0.118	0.120	0.119	0.118
Predictive goodness-of-fit measures in validation data (N=8,322)				
Predictive log-likelihood	-7427.6	-7410.9	-7416.4	-7415.4
Predictive AIC value	14897.1	14871.9	14878.8	14876.8

6.3. Empirical Findings

Table 5 reports the best-fitting empirical specification of the ML-ME-PLN model, which is the best-performing model of all the models estimated in this study. The findings from this model are discussed in detail, followed by a brief comparison with findings from the other models. The estimation results of other models are not reported in the table but are available in Nirmale (2022).

6.3.1. Empirical Findings on Perception Errors

Various empirical specifications were explored to incorporate stochasticity due to errors in perceiving the space gaps and relative speeds of the SV with respect to its surrounding vehicles. This includes: (1) a specification with each (and every) traffic environment variable having its own perception error term, (2) a specification with all space variables having a common error term and all relative speed variables having a common error term, and (3) the specification presented in this section, where perception error terms were specified to be common for all longitudinal space gaps with respect to vehicles in a given compartment (but different from those in other compartments); and similar specification for relative speed variables. As such, a total of ten PLN distributed stochastic terms were explored for the multiplicative error terms (τ_{qk}) in the model formulation. This specification provided the best fit and interpretation among all specifications.

Table 5 Estimation results of the ML-ME-PLN model *

Explanatory variables in the utility functions (maintain same speed is the base alternative)	**Acceleration utility**	**Deceleration utility**
Constant	1.970 (7.18)	-0.880 (-3.41)
Subject vehicle (SV) longitudinal speed (m/s)	-0.102 (-3.83)	0.231 (9.31)
Traffic environment variables with respect to MF1 (first vehicle in MF) at t-0.5 s		
Space gap in longitudinal direction (m)	0.023 (2.00)	-0.087 (-2.78)
Relative speed in longitudinal direction (m/s)	0.196 (4.53)	-0.259 (-5.28)
Traffic environment variables with respect to MF2 (second vehicle in MF) at t-0.5 s		
Subject vehicle has 2 or more lead vehicles (One lead vehicle is base)	-0.627 (-4.58)	--
Space gap in longitudinal direction (m)	0.021 (1.81)	--
Relative speed in longitudinal direction (m/s)	0.208 (3.70)	-0.120 (-2.60)
Traffic environment variables with respect to LF1 (first vehicle in LF) at t-0.5 s		
Subject vehicle has 1 or more lead vehicles (No lead vehicle is base)	--	0.448 (3.83)
Space gap in longitudinal direction (m)	--	-0.014 (-3.07)
Lateral gap between MF1 and LF1 (m)	0.109 (2.55)	--
Relative speed in longitudinal direction (m/s)	0.126 (2.40)	--
Traffic environment variables with respect to RF1 (first vehicle in RF) at t-0.5 s		
Subject vehicle has 1 or more lead vehicles (No lead vehicle is base)	--	--
Space gap in longitudinal direction (m)	--	--
Lateral gap between MF1 and RF1 (m)	--	--
Relative speed in longitudinal direction (m/s)	0.083 (3.21)	--
Traffic environment variables with respect to LS1 (first vehicle in LS) at t-0.5 s		
Subject vehicle has 1 or more side vehicle (No side vehicle is base)	-0.201 (-1.98)	--
Lateral space gap (m)	0.097 (3.29)	--
Relative speed in longitudinal direction (m/s)	--	--
Traffic environment variables with respect to RS1 (first vehicle in RS) at t-0.5 s		
Subject vehicle has 1 or more side vehicle (No side vehicle is base)	--	--
Lateral space gap (m)	--	--
Relative speed in longitudinal direction (m/s)	--	--
Position of subject vehicle (SV) at t-0.5 s		
Space gap between left edge of the SV and left edge of the road (m)	--	-0.121 (-6.55)
Variables on which perception error is considered in the ML-ME-PLN model	**Scale parameter**	**Standard deviation**
Longitudinal space gaps (m) - between SV & MF1 and between SV & MF2	2.501 (7.39)**	0.076
Relative longitudinal speeds (m/s) - between SV & MF1 and between SV & MF2	1.441 (5.38)**	0.160
Space gap (m) in longitudinal direction between SV & LF1	--	--
Relative speed (m/s) in longitudinal direction between SV & LF1	1.670 (1.87)***	0.743
Space gap (m) in longitudinal direction with respect to RF1	--	--
Relative speed (m/s) in longitudinal direction between SV & RF1	1.240 (1.21)***	0.598
Lateral gaps (m) between MF1 & LF1 and between MF1 & RF1	1.591 (1.66)***	0.715
Lateral gap (m) between SV & LS1 and between SV & RS1	--	--
Relative speed (m/s) in longitudinal direction between SV & LS1	--	--
Relative speed (m/s) in longitudinal direction between SV & RS1	--	--

Notes: *t-statistic for each estimated parameter is reported in parentheses next to it. ** Power value is fixed at 2.5. *** Power value is fixed at 1.5. -- the parameter was dropped from the specifications as it was insignificant. Maintain same speed is the base alternative.

The bottom set of rows in Table 5 reports the scale parameter estimates of the perception error distribution terms in the ML-ME-PLN model. As can be observed, the empirical model yielded stochasticity due to perception error in five sets of traffic environment variables: (a) longitudinal space gaps of the SV with respect to MF1 and MF2, (b) relative longitudinal speeds of the SV with respect to MF1 and MF2, (c) relative longitudinal speed of the SV with respect to LF1, (d) relative longitudinal speed of the SV with respect to RF1, and (e) lateral gaps between lead vehicles in the front compartments (i.e., MF1-LF1 lateral gap and MF1-RF1 lateral gap).

Statistically significant stochasticity was not uncovered for the other five sets of traffic environment variables due to different reasons. For example, we could not uncover a statistically significant scale parameter for the longitudinal space gap with respect to RF1. This result should not necessarily be interpreted as the absence of errors in perceiving this variable. To understand this, note from the earlier rows in the table that the variable does not enter the model specification. The coefficient of the variable was not statistically significant when the variable was included in the specification. The implication is that one cannot identify stochasticity due to perception error of a variable that does not have a strong influence on the choice outcome. For the same reason (that the variables did not have a statistically significant influence on the utility functions), the model did not yield statistically significant variability in the perception of relative speeds of the SV with respect to side vehicles (LS1 and RS1). For lateral gaps of the SV with respect to side vehicles (LS1 and RS1), stochasticity in perception was not uncovered, possibly because these vehicles tend to be in very close proximity to the SV, making its driver pay close attention to them.

Based on the scale and power parameters for the variables for which stochasticity in perception was uncovered, the standard deviation of the power lognormal distribution $\left(Std_{PLN}\right)$ may be calculated using the following expression:

$$Std_{PLN} = \sqrt{\left\{\int_0^1 \exp\left(-2\sigma\Phi^{-1}\left(y^{1/p}\right)+2\mu\right)dy\right\} - Mean^2} \tag{27}$$

The corresponding standard deviation parameters are reported in Table 5 in the column titled “Standard deviation.” Comparing the magnitudes of the standard deviation of relative speeds, it can be observed that the stochasticity due to perception error in relative speeds with respect to MF1 and MF2 is much lower than that for other vehicles (LF1, RF1). That is, a greater variation is reflected in driver perceptions of the traffic environment that is not directly ahead of them. This may be because drivers pay greater attention to vehicles directly ahead of their vehicle than those that are not ahead; hence, lower variability in perception for vehicles directly ahead of the SV. Another observation is that stochasticity due to perception errors for relative speeds with respect to MF1 and MF2 is greater than that for space gaps with respect to those vehicles. This result suggests that drivers perceive relative longitudinal speeds less precisely than longitudinal space gaps. Having said that, the drivers' perception of lateral gaps between two vehicles in the front compartments is associated with greater uncertainty in perception than that associated with longitudinal space gaps. This may be because perceiving a lateral gap between two moving vehicles is more difficult than perceiving a longitudinal gap with respect to one vehicle. While all these findings sound plausible, this is perhaps the first study to shed light on differences in the uncertainty of perception of different traffic environment variables. Therefore, additional empirical evidence is needed before stronger conclusions can be made.

6.3.2. Empirical Findings other than Perception Errors

The empirical model results in Table 5 offer interesting insights into driver behaviour. As expected, and reported by Koutsopoulos et al. (2012), a subject vehicle (SV) travelling at a higher longitudinal speed is less (more) likely to accelerate (decelerate) than that travelling at a slower speed. Further, as the relative speed or space gap of the SV with respect to the first lead vehicle (MF1) increases (decreases), the SV is more likely to accelerate (decelerate). In this context, the magnitude of the coefficient on space gap with respect to MF1 in the deceleration decision is greater than that for the acceleration decision. Similarly, the magnitude of the coefficient on relative speed with respect to MF1 in the deceleration decision is greater than that for the acceleration decision. These results suggest that the influence of space gap and relative speed is stronger on the decision to decelerate than on the decision to accelerate. This may be because the deceleration decision is more safety-critical than the acceleration decision.

As discussed in Nirmale et al. (2021), space gaps and relative speeds with respect to multiple vehicles influence the subject vehicle drivers' manoeuvring decisions highlighting the need to move beyond single-leader car-following models. In addition to the immediately leading and the next vehicle (MF1 and MF2) in the space directly ahead of the subject vehicle, vehicles in the left front (LF1), right front (RF1) and left side (LS1) also affect drivers' decisions. For each of the above vehicles, the parameter estimates are in line with the expected direction of their influence.

For example, greater (smaller) space gaps are associated with a higher (lower) likelihood of acceleration and greater relative speeds are associated with a higher likelihood of acceleration.

One of the differences between the proposed ML-ME-PLN model and the other models is worth noting. As per the ML-ME-PLN model, reducing the lateral gap between the vehicle in the MF1 and LF1 compartment might not increase the deceleration likelihood relative to the maintain same speed alternative. On the other hand, the MNL and the ML-RC-PLN models suggest otherwise, that reducing the lateral gap can lead to an increased likelihood of deceleration. This is perhaps because these models do not consider variability in the perception of the variable. In addition, several parameter estimates in the ML-ME-PLN (model with perception errors) have greater magnitudes than those in the MNL, ML-RC-PLN, and ML-CRC-PLN models – perhaps due to differences in the scales of the kernel error terms across the different models.

7. CONCLUSIONS

This paper proposes a discrete choice modelling framework that accommodates perception errors in choice environment variables that do not vary across choice alternatives. The framework takes the form of a mixed multinomial logit (ML) model where the choice environment variables under consideration are specified as stochastic. To operationalize this framework, an analysis is undertaken to evaluate two different ways of specifying stochastic variables in discrete choice models – (a) the additive EIV specification and (b) the multiplicative EIV specification. The additive specification assumes the errors to be independent of the magnitude of the quantity being perceived, whereas the multiplicative specification renders the variability in errors to depend on the magnitude of the quantity (i.e., higher errors for higher magnitudes). The latter is more suitable to represent errors in human perceptions of physical quantities. An econometric identification analysis suggests that only two variance-covariance parameters can be estimated for a discrete choice model with three alternatives with an additive error specification of errors in the choice environment variables. On the contrary, a multiplicative specification allows, in theory, separate identification of stochasticity in as many variables as the analyst would want to test – as long as the variables have a significant influence on the choice outcome. This helps in comparing the variability due to perception errors in different types of choice environment variables.

It is shown that the proposed model with multiplicative EIV for an attribute that does not change across alternatives is a special case of a more general model with correlated, alternative-specific random coefficients (CRC) on that attribute. However, if the primary source of stochasticity is due to multiplicative EIV and not due to random heterogeneity in the coefficients on the variable, then the more general CRC model cannot be estimated due to parameter (un)identifiability issues. In such a case, it is advisable to estimate the proposed multiplicative EIV model. Other typically used mixed discrete choice models, such as uncorrelated random coefficients or error components models (either for heteroskedasticity or inter-alternative correlations), are also not suitable in lieu of multiplicative perception errors. In addition to these conceptual discussions, we conducted extensive simulation experiments to verify this claim.

It is also shown, through simulation experiments, that when the underlying DGP is random coefficients on such choice environment variables (whether correlated or uncorrelated), the multiplicative EIV model does not provide better performance than the true DGP model unless the random coefficients have similar standard deviation values and high correlation. Of course, when both multiplicative EIV for a variable and correlated random heterogeneity in sensitivities to that variable are prevalent, it is difficult to separately identify the multiplicative EIV from correlated random coefficients. In such a case, it is preferable to estimate the CRC model.

We demonstrate the usefulness of the proposed multiplicative EIV model through an empirical application for analysing driver behaviour using space-time trajectories of vehicles from a traffic stream in Chennai, India. In this context, a subject vehicle's (SV) driver behaviour at an instance in a traffic stream is characterized as the driver's choice to accelerate, decelerate, or maintain same speed as a function of the various traffic environment variables such as space gaps and relative speeds between the SV and other vehicles around it. The empirical analysis results suggest that consistent with the findings from simulation experiments, the proposed ML model with power lognormal (PLN) distributed perception errors in traffic environment variables outperformed typically used ML models with random coefficients or error components. A correlated random coefficients (CRC) model showed signs of parameter unidentifiability and high correlation values suggesting that the primary source of stochasticity is due to errors in the traffic environment variables, not random coefficients on them. These results suggest that in driver behaviour models, it may be more important to accommodate drivers' errors in perceiving their traffic environment than to focus on random sensitivities to the traffic environment variables (if one must choose between the two). Of course, in other empirical contexts, one can always explore the correlated random coefficient model to explore the presence of both sources of stochasticity.

In the context of the distributional assumptions explored for perception errors, the PLN distribution allowed better estimability and offered a better fit to the empirical data than alternative distributions such as lognormal, Weibull, Rayleigh, exponential, and Fréchet.

The empirical model offered interesting insights on perception errors in traffic environment variables. First, greater variation was found in drivers' perceptions of the traffic environment variables with respect to vehicles that are not directly ahead of their vehicles (than those that are ahead). This may be because drivers pay greater attention to vehicles directly ahead of their vehicle than those that are not ahead. Second, stochasticity due to perception errors for relative longitudinal speeds was found to be greater than that for longitudinal space gaps – perhaps because drivers perceive relative speeds less precisely than space gaps. Third, drivers' perception of lateral gaps between two moving vehicles ahead is associated with greater uncertainty than that associated with longitudinal space gaps with respect to any of those vehicles. Fourth, it was not possible to recover variability due to perception errors for variables that did not influence the choice outcome.

The current study can be extended in different ways. First, the empirical analysis considers different observations of the same vehicle (and the corresponding utility functions) as independent from each other. However, correlations due to driver-specific unobserved factors that persist across all observations of the same vehicle have been ignored. Further, other aspects such as state-dependence, where a driver does not change the acceleration/deceleration/constant speed state unless there is a strong enough stimulus, ought to be considered. Also, the study does not consider the extent of acceleration or deceleration, which is an important dimension of driver behaviour. Such econometric and behavioural features ought to be explored in future work. Also, it would be useful to compare findings from different geographical and empirical contexts. While the findings from the current study sound plausible, this is perhaps the first study to shed light on differences in the uncertainty of perception of different traffic environment variables. Therefore, additional empirical evidence is needed before stronger conclusions can be made. Further, on the methodological front, much work needs to be done toward disentangling the stochasticity due to errors-in-variables and the stochasticity in random coefficients on those variables. Finally, the current study focuses on multiplicative stochasticity for errors in variables that do not vary across alternatives. There is scope to consider multiplicative stochasticity for perception errors in variables describing choice alternatives as well.

ACKNOWLEDGEMENTS

We benefited from discussions with Partha Chakroborty on the empirical model specification and the behavioural appeal of multiplicative specification for perception errors. We are thankful to Kanagaraj et al. (2015) for making the empirical data available to the research community. Two anonymous reviewers provided valuable comments on an earlier manuscript.

Appendix A. Estimation of the Proposed Mixed Logit Model with Multiplicative Errors

The parameters of the proposed model can be estimated using the MSL approach. To do so, building on Eq. (3) for the likelihood expression for a driver q's manoeuvring choice, the likelihood expression for a sample of independent observations $(q=1,2,...,Q)$ may be written as:

$$L=\prod_{q=1}^{Q} L_q(\beta,\theta)=\prod_{q=1}^{Q}\prod_{i=a,d,s}\left\{L_{qi}\left(\beta,\theta\right)\right\}^{\delta_{qi}} \tag{A.1}$$

where, $\delta_{qi}=1$ if the driver of vehicle q chooses manoeuvring alternative i; zero otherwise. Considering multiplicative perception errors, the individual likelihood term $L_{qi}\left(\beta,\theta\right)$ in the above expression may be written as below:

$$L_{qi}\left(\beta,\theta\right)=\int_{\tau_q} L_{qi}\left(\beta,x_q,\tau_q\right) f\left(\tau_q\right) d\tau_q \tag{A.2}$$

where, τ_q is a vector of all perception error terms τ_{qk} $(k=1,2,...,K)$ corresponding to x_q, which is in turn a vector of measurements of choice environment variables x_{qk} $(k=1,2,...,K)$. And $L_{qi}\left(\beta,x_q,\tau_q\right)$ is the conditional likelihood function (conditioned on the values of τ_q) that the driver of a vehicle q makes a manoeuvring choice i, given by the following expression:

$$L_{qi}\left(\beta,x_q,\tau_q\right)=\frac{\exp\left(\beta_{i0}+\sum_{k=1}^{K}\beta_{ik}x_{qk}\tau_{qk}\right)}{\sum_{j=a,d,s}\exp\left(\beta_{j0}+\sum_{k=1}^{K}\beta_{jk}x_{qk}\tau_{qk}\right)} \tag{A.3}$$

The multivariate integral in the likelihood function of Eq. (A.2) may be simulated to result in the following simulated likelihood function as an estimator of $L_{qi}\left(\beta,\theta\right)$:

$$SL_{qi}(\beta,\theta)=\frac{1}{R}\sum_{r=1}^{R}\frac{\exp\left(\beta_{i0}+\sum_{k=1}^{K}\beta_{ik}x_{qk}\tau_{qk}^{r}\right)}{\sum_{j=a,d,s}\exp\left(\beta_{j0}+\sum_{k=1}^{K}\beta_{jk}x_{qk}\tau_{qk}^{r}\right)} \tag{A.4}$$

where, τ_q^r is the r^{th} draw from the distribution of the vector τ_q of perception error terms and R is the total number of such draws covering the distribution of τ_q. The corresponding simulated log-likelihood function $SLL(\beta,\theta)$ for the entire data is given in the expression below:

$$SLL(\beta,\theta)=\sum_{q=1}^{Q}\sum_{i=a,d,s}\delta_{qi}\ln\left\{SL_{qi}(\beta,\theta)\right\} \tag{A.5}$$

To estimate parameters using the MSL method, we apply quasi-Monte Carlo simulation techniques to draw from the distribution of τ_q for simulating $SLL(\beta,\theta)$. Specifically, 400 sets of Halton draws (i.e., $R=400$) were used to simulate τ_q. For the distributions explored for τ_q in this study, it was not easy to achieve convergence of the MSL parameter estimation routine when numerically computed gradients were used. Therefore, analytical gradients of the simulated log-likelihood function were also coded to assist in estimation. Next, we present the expressions for analytical gradients of the function $\ln\left\{SL_{qi}(\beta,\theta)\right\}$.

For simplicity in notation, denote the simulated likelihood $SL_{qi}(\beta,\theta)$ for the driver of a vehicle q choosing a manoeuvring alternative i as SL_{qi}. Also, rewrite Eq. (A.4) as, $SL_{qi}=\frac{1}{R}\sum_{r=1}^{R}L_{qi}^{r}$, where, L_{qi}^{r} is the likelihood function value computed at the r^{th} draw of τ_{qk} (i.e., at τ_{qk}^{r}); written as: $L_{qi}^{r}=\frac{\exp\left(\sum_{k=1}^{K}\beta_{ik}\,x_{qk}\tau_{qk}^{r}\right)}{\sum_{j=1}^{J}\exp\left(\sum_{k=1}^{K}\beta_{jk}\,x_{qk}\tau_{qk}^{r}\right)}$.

Using the above notation, the gradient with respect to β_{jk} of the simulated log-likelihood $\ln\left\{SL_{qi}(\beta,\theta)\right\}$ may be derived as below (derivation details are available with the authors):

$$G_{qi}\left(\beta_{jk}\right)=\frac{1}{SL_{qi}}\left[\frac{1}{R}\sum_{r=1}^{R}\left\{L_{qi}^{r}\times\sum_{\forall l\in C_q:\beta_{lk}=\beta_{jk}}x_{qk}\tau_{qk}^{r}\left(\partial_{ql}-L_{ql}^{r}\right)\right\}\right] \tag{A.6}$$

In the above expression, l is an index for choice alternatives. C_q is the choice set for individual driver q. Note that the summation $\sum_{\forall l \in C_q : \beta_{lk} = \beta_{jk}} (.)$ in the above expression is useful when a coefficient β_{jk} is specified to be same across a subset of choice alternatives (although it was not necessary to do so in this study). If β_{jk} is specific to only the j^{th} alternative, then the only term in the summation corresponds to the j^{th} alternative. And the indicator variable $\partial_{ql} = 1$ if $l = i$; zero otherwise. That is, ∂_{ql} takes a value 1 when the gradient is being taken with respect to a parameter of the chosen alternative. It should also be noted that when k^{th} variable is not a stochastic variable, then $\tau_{qk}^{r} = 1$ (i.e., $x_{qk}^{*} = x_{qk}$).

The gradient with respect to σ_k of the simulated log-likelihood is derived as:

$$G_{qi}(\sigma_k) = \frac{x_{qk}}{SL_{qi}} \left[\frac{1}{R} \sum_{r=1}^{R} \left\{ L_{qi}^{r} \frac{\partial \tau_{qk}^{r}}{\partial \sigma_k} \left(\beta_{ik} - \sum_{j \in C_q'} \beta_{jk} L_{qj}^{r} \right) \right\} \right] \tag{A.7}$$

Here, C_q' denotes the subset of choice alternatives for which x_{qk}^{*} enters the utility function. $\frac{\partial \tau_{qk}^{r}}{\partial \sigma_k}$ is a partial derivative of the inverse CDF function of the distribution assumed for τ_{qk}. Table A.1 provides expressions of this partial derivative for the distributions explored in this study when their location parameters are set such that the expected value of the distribution is 1.

Table A.1 Partial derivative of τ_{qk}^r with respect to σ_k for different distributions of τ_{qk}

Distribution of τ_{qk}	$\frac{\partial \tau_{qk}^r}{\partial \sigma_k}$
Power lognormal	$\tau_{qk}^r \left[-\Phi^{-1}\left[(1-u^r)^{1/p}\right] + \frac{\partial}{\partial \sigma_k} - \ln\left\{ \int_0^1 \exp\left(-\sigma_k \Phi^{-1}\left[y^{1/p}\right]\right) dy \right\} \right]$
Lognormal	$\tau_{qk}^r \left(-\sigma_k + \Phi^{-1}\left[u^r\right]\right)$
Weibull	$\left[-\ln(1-u^r)\right]^{1/\gamma} - \Gamma(\gamma_k^{-1}+1)$
Rayleigh	$\sqrt{-2\ln(1-u^r)} - \sqrt{\frac{\pi}{2}}$
Exponential	$\ln\left(1-u^r\right) - 1$
Fréchet	$\frac{-\ln u^r}{\alpha_k} - \Gamma\left(1 - \frac{1}{\alpha_k}\right)$

Notes: In the above expressions, u^r is r^{th} draw from a uniform [0,1] distribution. $\Phi^{-1}[.]$ is inverse CDF function of standard normal distribution. To compute the expression $\frac{\partial}{\partial \sigma_k} - \ln\left\{ \int_0^1 \exp\left(-\sigma_k \Phi^{-1}\left[y^{1/p}\right]\right) dy \right\}$ corresponding to the power lognormal distribution, the integral inside it and, subsequently, the partial derivative was computed numerically. For the Weibull distribution, partial derivative of τ_{qk}^r is with respect to κ_k, not with respect to σ_k.

Appendix B. Alternative Distributions Explored for Representing Drivers' Perception Errors

Table B.1 Alternative distributions explored for representing drivers' perception errors

Distribution name	**Density function** $f(Z)$	**Location parameter** μ **when** $E[Z]=1$	**Inverse CDF function when** $E[Z]=1$	**Standard deviation when** $E[Z]=1$	**Notes**
Power lognormal	$\left(\frac{p}{Z\sigma}\right)\phi\left(\frac{\ln Z-\mu}{\sigma}\right)\left\{\Phi\left[-\left(\frac{\ln Z-\mu}{\sigma}\right)\right]\right\}^{p-1}$	$-\ln\left(\int_0^1 \exp(-\sigma\Phi^{-1}(y^{1/p}))dy\right)$	$\exp\left(-\sigma\Phi^{-1}[(1-u)^{1/p}]-\ln\left(\int_0^1 \exp(-\sigma\Phi^{-1}(y^{1/p}))dy\right)\right)$	$\sqrt{\left\{\int_0^1 \exp\left(-2\sigma\Phi^{-1}(y^{1/p})+2\mu\right)dy\right\}-1}$	$\sigma>0,-\infty<\mu<\infty,p>0$ Support:$(0,\infty)$
Lognormal	$\frac{1}{Z\sigma}\phi\left(\frac{\ln Z-\mu}{\sigma}\right)$	$-\frac{\sigma^2}{2}$	$\exp\left(\frac{-\sigma^2}{2}+\sigma\Phi^{-1}(u)\right)$	$\exp\left(-\frac{\sigma^2}{2}\right)\sqrt{\exp\left(\sigma^2\right)\left[\exp\left(\sigma^2\right)-1\right]}$	$\sigma>0,-\infty<\mu<\infty$, Support:$(0,\infty)$
Weibull	$\left(\frac{\gamma}{\kappa}\right)\left(\frac{Z-\mu}{\kappa}\right)^{\gamma-1}\exp\left[-\left(\frac{Z-\mu}{\kappa}\right)^{\gamma}\right]$	$1-\kappa\Gamma(\gamma^{-1}+1)$	$\kappa\left[-\ln(1-u)\right]^{1/\gamma}+1-\kappa\Gamma(\gamma^{-1}+1)$	$\kappa\left[\Gamma(1+2\gamma^{-1})-\left\{\Gamma(1+\gamma^{-1})\right\}^2\right]$	$\kappa>0,\mu\geq 0,\gamma>0$ Support:$[\mu,\infty)$
Rayleigh	$\left(\frac{Z-\mu}{\sigma^2}\right)\exp\left[-\frac{1}{2}\left(\frac{\ln Z-\mu}{\sigma}\right)^2\right]$	$1-\sigma\sqrt{\frac{\pi}{2}}$	$\sigma\sqrt{-2\ln(1-u)}+1-\sigma\sqrt{\frac{\pi}{2}}$	$\sigma\sqrt{\frac{4-\pi}{2}}$	$\sigma>0,\mu\geq 0$, Support:$[\mu,\infty)$
Exponential	$\frac{1}{\sigma}\exp\left[-\left(\frac{\ln Z-\mu}{\sigma}\right)^2\right]$	$1-\sigma$	$-\sigma\ln(1-u)+1-\sigma$	σ	$\sigma>0,\mu\geq 0$, Support:$[\mu,\infty)$
Fréchet	$\frac{\alpha}{\sigma}\left(\frac{Z-\mu}{\sigma}\right)^{-1-\alpha}\exp\left(-\frac{Z-\mu}{\sigma}\right)^{-\alpha}$	$1-\sigma\Gamma\left(1-\frac{1}{\alpha}\right)$	$\frac{-\sigma\ln u}{\alpha}+1-\sigma\Gamma\left(1-\frac{1}{\alpha}\right)$	$\begin{cases}\sigma^2\left[\Gamma\left(1-\frac{2}{\alpha}\right)-\left(\Gamma\left(1-\frac{1}{\alpha}\right)\right)^2\right] & \text{for } \alpha>2 \\ \infty & \textit{otherwise}\end{cases}$	$\sigma>0,-\infty<\mu<\infty,\alpha>0$ Support:$[\mu,\infty)$

Notes: This is a modified version of a similar table provided in Bhat and Lavieri (2018). Lognormal distribution is a special case of power lognormal distribution when the latter's shape parameter p is set to 1. Weibull distribution collapses to exponential distribution when the former's shape parameter (γ) is equal to 1. Weibull collapses to Rayleigh when its shape parameter $\gamma=2$ and scale parameter (κ) is equal to $\sqrt{2}\sigma$. Shape parameter of Fréchet distribution is denoted by α. Support for the power lognormal distribution (and lognormal distribution) is the strictly positive semifinite interval (0,∞). Support for Weibull, Rayleigh, Exponential, and Frechet distributions is [μ,∞), where μ is the minimum value.

REFERENCES

Ahmed, K. I. (1999). Modeling drivers' acceleration and lane changing behavior. *PhD thesis*, Massachusetts Institute of Technology.

Allan, L. G. (2001). *International Encyclopedia of Social & Behavioral Sciences*, 15696–15699.

Bartels, R., Fiebig, D. G., and Van Soest, A. (2006). Consumers and experts: an econometric analysis of the demand for water heaters. *Empirical Economics*, *31*(2), 369-391.

Bevrani, K., and Chung, E. (2012). A safety adapted car following model for traffic safety studies. *Advances in Human Aspects of Road and Rail Transportation*, 550-559.

Bhat, C. R., and Lavieri, P. S. (2018). A new mixed MNP model accommodating a variety of dependent non-normal coefficient distributions. *Theory and Decision*, *84*(2), 239-275.

Bhatta, B. P., and 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.

Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

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

Bolduc, D., and Alvarez-Daziano, R. (2010). On estimation of hybrid choice models. *Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference*, 259-287.

Bunch, D. S. (1991). Estimability in the multinomial probit model. *Transportation Research Part B: Methodological*, *25*(1), 1-12.

Cardell, N. S., and Dunbar, F. C. (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research Part A: General*, *14*(5), 423-434.

Carroll, R. J., Spiegelman, C. H., Lan, K. G., Bailey, K. T., and Abbott, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika*, *71*(1), 19-25.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC.

Castillo, E., Menéndez, J. M., Jiménez, P., and Rivas, A. (2008). Closed form expressions for choice probabilities in the Weibull case. *Transportation Research Part B: Methodological*, *42*(4), 373-380.

Chakroborty, P., Pinjari, A. R., Meena, J., and Gandhi, A. (2021). A Psychophysical Ordered Response Model of Time Perception and Service Quality: Application to Level of Service Analysis at Toll Plazas. *Transportation Research Part B: Methodological*, *154*, 44-64.

Chikaraishi, M., and Nakayama, S. (2016). Discrete choice models with q-product random utilities. *Transportation Research Part B: Methodological*, *93*, 576-595.

Choudhury, C. F. (2007). Modeling driving decisions with latent plans. *PhD thesis,* Massachusetts Institute of Technology.

Choudhury, C. F., and Islam, M. M. (2016). Modelling acceleration decisions in traffic streams with weak lane discipline: A latent leader approach. *Transportation Research Part C: Emerging Technologies*, *67*, 214-226.

Daganzo, C. F., and Sheffi, Y. (1977). On stochastic models of traffic assignment. *Transportation science*, *11*(3), 253-274.

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

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

Fechner, G. T., Howes, D. H., and Boring, E. G. (1966). Elements of psychophysics. Holt, Rinehart and Winston, New York.

Fosgerau, M., and Bierlaire, M. (2009). Discrete choice models with multiplicative error terms. *Transportation Research Part B: Methodological*, *43*(5), 494-505.
Fuller, W. A. (2009). Measurement error models. John Wiley & Sons.
Gipps, P. G. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, *15*(2), 105-111.
Gray, R., and Regan, D. (1998). Accuracy of estimating time to collision using binocular and monocular information. *Vision Research*, *38*(4).
Greene, W. H. (2018). Econometric analysis. Pearson Education India.
Hamdar, S. (2012). Driver Behavior Modeling. *Handbook of Intelligent Vehicles*, 537-558.
Hamdar, S., Mahmassani, H. S., and Treiber, M. (2015). From behavioral psychology to acceleration modeling: Calibration, validation, and exploration of drivers' cognitive and safety parameters in a risk-taking environment. *Transportation Research Part B: Methodological*, *78*, 32-53.
Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
Hellerstein, D. (2005). Modeling Discrete Choice with Uncertain Data: An Augmented MNL Estimator. *American Journal of Agricultural Economics*, *87*(1), 77-84.
Hess, S., Daly, A., Dekker, T., Cabral, M. O., and Batley, R. (2017). A framework for capturing heterogeneity, heteroskedasticity, non-linearity, reference dependence and design artefacts in value of time research. *Transportation Research Part B: Methodological*, *96*, 126-149.
Hess, S., and Train, K. (2017). Correlation and scale in mixed logit models. *Journal of choice modelling*, *23*, 1-8.
Hoogendoorn, S., Hoogendoorn, R., and Daamen, W. (2011). Wiedemann Revisited:New Trajectory Filtering Technique and Its Implications for Car-Following Modeling. *2260*(1), 152-162.
Horowitz, J. L. (1983). Statistical comparison of non-nested probabilistic discrete choice models. *Transportation Science*, *17*(3), 319-350.
Kanagaraj, V., Asaithambi, G., Toledo, T., and Tzu-Chang, L. (2015). Trajectory data and flow characteristics of mixed traffic. *Transportation Research Record: Journal of the Transportation Research Board*, *2491*(1), 1-11.
Kao, C., and Schnell, J. F. (1987). Errors in variables in the multinomial response model. *Economics Letters*, *25*(3), 249-254.
Kikuchi, S., and Chakroborty, P. (1992). Car-following model based on fuzzy inference system. *Transportation Research Record*, 82-82.
Koutsopoulos, H. N., and Farah, H. (2012). Latent class model for car following behavior. *Transportation Research Part B: Methodological*, *46*(5), 563-578.
Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3), 229-254.
McFadden, D. (1984). Econometric analysis of qualitative response models. *Handbook of Econometrics*, *2*, 1395-1457.
McFadden, D., and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, *15*(5), 447-470.
Nirmale, S. K., Pinjari, A. R., and Sharma, A. (2021). A discrete-continuous multi-vehicle anticipation model of driving behaviour in heterogeneous disordered traffic conditions. *Transportation Research Part C: Emerging Technologies*, *128*.
Nirmale, S. K. (2022). Multi-Vehicle Anticipation-based Models for Describing Driver Behaviour in Heterogeneous and Disorderly Traffic Conditions. *PhD thesis*, Indian Institute of Science.
Ojeda-Cabral, M., Batley, R., and Hess, S. (2016). The value of travel time: random utility versus random valuation. *Transportmetrica A: Transport Science*, *12*(3), 230-248.
Ojeda-Cabral, M., and Chorus, C. G. (2016). Value of travel time changes: Theory and simulation to understand the connection between Random Valuation and Random Utility methods. *Transport Policy*, *48*, 139-145.

Ortúzar, J. d. D., and Ivelic, A. (1987). Effects of using more accurately measured level-of-service variables on the specification and stability of mode choice models. *Proceeding 15th PTRC Summer Annual Meeting*, *290*, 117-130.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. John Wiley & Sons.

Saifuzzaman, M., and Zheng, Z. (2014). Incorporating human-factors in car-following models: a review of recent developments and research needs. *Transportation research part C: Emerging Technologies*, *48*, 379-403.

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

Sharma, A., Ali, Y., Saifuzzaman, M., Zheng, Z., and Haque, M. M. (2018). Human Factors in Modelling Mixed Traffic of Traditional, Connected, and Automated Vehicles. Cham, 262-273.

Stefanski, L. A., and Carroll, R. J. (1985). Covariate Measurement Error in Logistic Regression. *The Annals of Statistics*, *13*(4), 1335-1351.

Steinmetz, S. S., and Brownstone, D. (2005). Estimating commuters'"value of time" with noisy data: a multiple imputation approach. *Transportation Research Part B: Methodological*, *39*(10), 865-889.

Toledo, T. (2003). Integrated driving behavior modeling. *PhD thesis*, Massachusetts Institute of Technology.

Train, K. E. (1978). The sensitivity of parameter estimates to data specification in mode choice models. *Transportation*, *7*(3), 301-309.

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge university press.

Treiber, M., Kesting, A., and Helbing, D. (2006). Delays, inaccuracies and anticipation in microscopic traffic models. *Physica A: Statistical Mechanics and its Applications*, *360*(1), 71-88.

Treiber, M., and Kesting, A. (2013). Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg*.

van Lint, H., Calvert, S., Schakel, W., Wang, M., and Verbraeck, A. (2017). Exploring the effects of perception errors and anticipation strategies on traffic accidents-A simulation study. *AHFE 2017*, 249-261.

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

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

Walker, J. (2001). Extended discrete choice models: integrated framework, flexible error structures, and latent variables. *PhD thesis*, Massachusetts Institute of Technology.

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

Wansbeek, T., and Meijer, E. (2000). Measurement Error and Latent Variables in Econometrics. *Economics Letters*, *69*.

Wiedemann, R. (1974). Simulation des StraBenverkehrsflusses. *Proceedings of the Schriftenreihe des tnstituts fir Verkehrswesen der Universitiit Karlsruhe, Germany*.

Wooldridge, J. M. (2012). Introductory econometrics: A modern approach. South-Western, Cengage Learning.

Yang, H. H., and Peng, H. (2010). Development of an errorable car-following driver model. *Vehicle System Dynamics*, *48*(6), 751-773.

Yatchew, A., and Griliches, Z. (1985). Specification Error in Probit Models. *The Review of Economics and Statistics*, *67*(1), 134-139.