MULTI-VEHICLE ANTICIPATION-BASED MODELS FOR DESCRIBING DRIVER BEHAVIOUR IN HETEROGENEOUS AND DISORDERLY TRAFFIC CONDITIONS

A THESIS

SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE FACULTY OF ENGINEERING

By

Sangram Krishna Nirmale

Department of Civil Engineering

Indian Institute of Science

Bengaluru – 560012

December 2022

DEDICATED

TO

My Parents and Teachers

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor Prof. Abdul Rawoof Pinjari. His innovative ideas served as the basis for this research, and his encouragement, meticulousness, and quest for excellence significantly improved the quality of the thesis. Working with him has been an honour and a life-changing experience. In addition, I want to express my gratitude to him for each and every way he has supported me on both personal and professional fronts.

Thanks to Dr. Anshuman Sharma, who has been a friend, guide and collaborator during my doctoral study. His technical and practical insights were invaluable for my research work. I also got the opportunity to collaborate with an eminent academic and researcher, Prof. Partha Chakroborty. I would like to thank him for his guidance, motivation, and insights during our collaboration. I want to thank Dr. Caleb Ronald Munigety, who introduced me to driver behaviour modelling. Without him, I would probably not have started my research work on modelling driver behaviour. I am also grateful to Kanagaraj et al. (2015), Punzo et al. (2011), Montanino and Punzo (2013, 2015) for making vehicular trajectory data available to the research community, which I used extensively in this research.

Prof. Ashish Verma and Prof. Tarun Rambha provided many helpful suggestions from different perspectives during my PhD. I thank them for their advice and encouragement. Also, many thanks to my former advisor Prof. Vidya Ghuge for encouragement and advice over the past several years. I am also thankful to each and everyone in the Department of Civil Engineering and CiSTUP at IISc for all the support during my PhD journey.

I was extremely fortunate to have an amazing group of friends who eased stressful times and extended their warmth – Shobhit, Mehek, Deepa, Vishal, Prateek, Rito, Helen, and other peers at IISc – thanks for everything. Many thanks to Shobhit and Prateek on many personal and professional fronts.

I could not have reached this point in my life without the care, unending support and guidance of my family. My deepest appreciation goes to my parents, Prof. Krishna Narasu Nirmale and Rukmini Nirmale, for instilling in me countless life principles, many of which I have found to be incredibly valuable during this journey. Thanks to my sister Dr. Sarvada and my brother Samruddh for their unconditional love, encouragement and unwavering support in all my endeavours.

Lastly, a special thanks to Namrata, my wife. She stood by me through the final and difficult part of my doctoral study. I could focus on my thesis because she was present in my absences. In this journey, I understood how important is her love, affection, patience and helping hand she gave me round the clock when I needed to discuss ideas.

DECLARATION

I, **Sangram Krishna Nirmale**, with SR No. 05-05-00-10-12-16-1-13947 hereby declare that the material presented in this thesis titled

Multi-Vehicle Anticipation-Based Models for Describing Driver Behaviour in Heterogeneous and Disorderly Traffic Conditions

is my own, except where the work is part of a collaboration, which is explicitly acknowledged.

Sangram Krishna Nirmale

PUBLICATIONS/CONFERENCE PRESENTATIONS FROM THIS THESIS

Journal Articles - Published/Accepted

1. **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. https://doi.org/10.1016/j.trc.2021.103144
2. **Nirmale, S. K.**, Pinjari, A. R., and Sharma, A. (2022). A panel data-based discrete-continuous modelling framework to analyse longitudinal driver behaviour in homogeneous and heterogeneous disordered traffic conditions. *Transportation Letters*. https://doi.org/10.1080/19427867.2022.2132058

Recently Communicated Papers in Journals

1. **Nirmale, S. K.**, and Pinjari, A. R. (2022). Discrete choice models with multiplicative stochasticity in choice environment variables: application for accommodating perception errors in driver behaviour models. *(In second round review with Transportation Research Part B: Methodological)*
2. **Nirmale, S. K.**, Sharma, A., and Pinjari, A. R. (2022). Multi-vehicle anticipation based driver behaviour models: A synthesis of existing research and future research directions. *(In review with Transportation Letters)*
3. **Nirmale, S. K.**, Pinjari, A. R., and Chakroborty P. (2022). A Two-Dimensional, Multi-Vehicle Anticipation, and Multi-Stimuli Based Latent Class Framework to Model Driver Behaviour in Heterogeneous, Disorderly Traffic Conditions. *(In review with Transportation Research Part C: Emerging Technologies)*

Papers Presented at Conferences

1. **Nirmale, S. K.**, Pinjari, A. R., and Sharma, A. (2022). On the capability of random utility maximisation theory-based driver behaviour models of heterogeneous disordered traffic to reproduce the fundamental macroscopic properties of traffic flow. At Traffic and Granular Flow (TGF), Indian Institute of Technology Delhi, India.
2. **Nirmale, S. K.**, Pinjari, A. R., and Chakroborty, P. (2022). Multi-vehicle anticipation based discrete-continuous choice modelling framework to model drivers' latent intents and two-dimensional movement in heterogeneous disordered traffic conditions. At the 7th International Choice Modelling Conference (ICMC) 2022, Reykjavík, Iceland.

3. **Nirmale, S. K.**, and Pinjari, A. R. (2020). Discrete-Continuous Choice Framework to Model Driver Behaviour in Heterogeneous Traffic Conditions. At the 12th International Conference on COMmunication Systems & NETworkS (COMSNETS) 2020, Bengaluru, India.
4. **Nirmale, S. K.**, and Pinjari, A. R. (2020). Driver Behaviour Models with Perception Errors: A Choice Modelling Framework with Stochastic Variables. At the Annual Meeting of the Transportation Research Board (TRB) 2020, Washington, DC, United States.
5. **Nirmale, S. K.** and Pinjari, A. R. (2019). Multi-Stimuli Driver Behaviour Models with Perception Errors: An Integrated Latent Variable and Discrete-Continuous Framework with Empirical Applications to Heterogeneous and Homogeneous Traffic Conditions. At the 6th International Choice Modelling Conference (ICMC) 2019, Kobe, Japan.
6. **Nirmale, S. K.**, Pinjari, A. R., and Munigety, C. R. (2019). A copula-based joint multinomial discrete-continuous choice framework to model driver behaviour in mixed traffic conditions. At the 15th World Conference on Transport Research (WCTR) 2019, Indian Institute of Technology, Mumbai, India.
7. **Nirmale, S. K.**, and Pinjari, A. R. (2018). Influence Zone, Multi-Stimuli, and Two-Dimensional (IZMS-2D) Driving Behaviour in Heterogenous Traffic Conditions: An Econometric Framework and Exploratory Analysis of Driving Behaviours in India. At the 15th International Conference on Travel Behavior Research (IATBR) 2018, Santa Barbara, California, United States.

Award

1. The research paper 'Driver Behaviour Models with Perception Errors: A choice Modelling Framework with Stochastic Variables' by **Sangram Krishna Nirmale** and Abdul Rawoof Pinjari was adjudged the best paper in the area of 'Traffic Flow Analysis and Facilities Design' at the 5th Conference of Transportation Research Group of India (CTRG) 2019.

ABSTRACT

Multi-Vehicle Anticipation-Based Models for Describing Driver Behaviour in Heterogeneous and Disorderly Traffic Conditions

by

Sangram Krishna Nirmale

Department of Civil Engineering, Indian Institute of Science, Bengaluru

Supervisor: Prof. Abdul Rawoof Pinjari

Driver behaviour models are widely used in the traffic engineering literature and practice. They are used for understanding drivers' manoeuvring decisions in traffic streams. They also form the building blocks of microscopic traffic simulation tools, which are employed for traffic flow analysis and capacity estimation necessary for the design and operation of traffic facilities and evaluation of operational strategies. Most driver behaviour models in the literature assume homogeneous and orderly traffic conditions, characterised by homogeneity (i.e., only passenger cars comprise the traffic streams) and orderliness (i.e., vehicles move only in the longitudinal direction, except when changing lanes). Models developed with such assumptions cannot be applied to analyse *heterogeneous, disorderly (HD) traffic* conditions. This is because HD traffic streams, unlike *homogeneous traffic* streams, comprise a wide variety of vehicle classes with considerably different physical and operational characteristics. Moreover, driving in HD traffic streams is characterised by weaker lane discipline due to a greater extent of lateral movements than that in *homogeneous traffic* streams.

This dissertation aims to formulate and apply driver behaviour models for HD traffic streams on uninterrupted traffic facilities while considering the following aspects – (1) the multi-vehicle anticipation (MVA) behaviour, where drivers' manoeuvring decisions are influenced by multiple vehicles around them, as opposed to a single lead vehicle ahead, (2) the treatment of driver behaviour as a combination of different manoeuvring decisions, such as the decision of whether to accelerate, decelerate, or remain in same speed (represented by a discrete variable) and the decision of the extent of acceleration or deceleration (represented by continuous variables) – as opposed to a single, continuous variable representing all these facets of driver behaviour, (3) the incorporation of stochasticity due to the errors drivers make in perceiving the traffic environment, and (4) the consideration of drivers' intentions (which are

typically latent to the analyst) and two-dimensional movements of vehicles simultaneously while also incorporating MVA behaviour. Specifically, the following driver behaviour models are formulated and applied to understand driver behaviour in empirical trajectory datasets from Chennai (HD traffic) and California (*homogeneous traffic*):

1. The first model presented in this dissertation is an MVA-based discrete-continuous choice modelling framework to model vehicles' longitudinal movements in HD traffic streams. In this model, driver behaviour at a given time instance is treated as a combination of (a) the driver's choice of whether to accelerate, decelerate, or maintain a constant speed – represented by a discrete variable – and (b) the extent of acceleration or deceleration – represented by continuous variables. The discrete and continuous variables representing driver behaviour are modelled using a simultaneous econometric framework. The proposed model is used to examine driver behaviour in the HD traffic dataset from Chennai. The empirical analysis reveals the significance of the MVA effect on driver behaviour. Specifically, drivers consider the relative speeds and space gaps with respect to multiple vehicles within an influence zone around their vehicle. In addition, it is found that the influence of the traffic environment on drivers' discrete choices (whether to accelerate, decelerate, or maintain a constant speed) is not the same as that on their choices of how much to accelerate or decelerate.

2. The second model is an extension of the above model to recognise the panel data nature of vehicle trajectory datasets typically used for estimating the parameters of driver behaviour models. This model recognises the role of vehicle- and driver-specific unobserved factors (latent to the analyst), such as aggressiveness that influence driving behaviour, and such influence persists across all observations of a vehicle. Doing so helps in reducing the confounding effects of unobserved factors when the proposed model is applied to different datasets to compare driving behaviour in different traffic streams. The panel data model is used to understand and compare longitudinal driving behaviour between the HD traffic dataset of Chennai and the *homogeneous traffic* dataset of California. The empirical analysis reveals the presence of MVA effect on driving behaviour in the homogeneous traffic setting, too. However, drivers in the HD traffic stream are influenced by more vehicles in their vicinity than those in the *homogeneous traffic* stream.

3. In the third model formulation, a mixed multinomial logit-based framework is developed to recognise stochasticity in driver behaviour models due to drivers' errors in perceiving

the traffic environment. For this model, an econometric analysis is undertaken to evaluate two different ways of specifying errors in variables in discrete choice models – additive errors and multiplicative errors. It is shown that the multiplicative specification of errors has a better behavioural basis and allows better identification of parameters representing variability due to drivers' perception errors. An application of this model to the HD traffic dataset reveals different levels of variability due to errors in the perception of different traffic environment variables. It is found that drivers may pay greater attention to (which results in lower variability in) perceiving space gaps and relative speeds with respect to vehicles directly ahead of them than those not directly ahead.

4. The fourth and final model formulation is a two-dimensional, MVA, and multi-stimuli-based latent class framework to analyse motorcyclists' two-dimensional movements in HD traffic streams. This formulation conjectures that drivers manage their cognitive load by dividing their driving decisions into two steps – (a) higher-level, strategic intentions (of whether to accelerate, decelerate, or maintain a constant speed and whether to steer to the left of, right of, or keep straight along the longitudinal direction), which are not fully observable from vehicle trajectories (hence latent to the analyst), and (b) lower-level, tactical decisions that can be observed in vehicle trajectories, such as the specific angle of movement and the specific extent of acceleration or deceleration executed. When applied to the HD traffic dataset of Chennai, the proposed model suggests that drivers' higher-level intentions are more strongly influenced by the microscopic traffic environment variables than their lower-level decisions, perhaps because drivers invest a greater extent of cognitive resources for making higher-level intentions than that for lower-level decisions.

Finally, a traffic simulator is developed to simulate traffic streams using the models developed in this dissertation. The simulation experiments using this simulator demonstrate that all the microscopic driver behaviour models developed in this dissertation reflect the typically observed macroscopic properties of vehicular traffic steams.

Keywords: heterogeneous disorderly traffic conditions, homogeneous traffic conditions, driver behaviour, multi-vehicle anticipation, discrete-continuous models, perception error, two-dimensional movements

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND NOTATIONS

MVA	Multiple-vehicle anticipation
HD	Heterogeneous and disorderly
2D	Two-dimensional
GHR	Gazis-Herman-Rothery
OVM	Optimal velocity model
FVD	Full velocity difference
IDM	Intelligent driver model
α_j, β_j, γ_j, α_1, α_2, α_3, λ_0, λ_1, λ_2, p_j, q_j	Sensitivity or weighting coefficients
V_1, V_2, C, C_0, C_1, C_2	Constants
λ	Parameter
$a_i(t)$	Acceleration/deceleration value of subject vehicle i at time t
$a_i^{\max}$	Maximum acceleration of vehicle i
b	Braking deceleration
b_i^{conf}	Comfortable deceleration of vehicle i
M	Mass of the vehicle
N, N_1, N_2	Number of lead vehicles
i	Subject vehicle index
j	Lead vehicle index
t	Time instance
$V_i(t)$	Velocity of vehicle i at time instance t (m/s)

V_{equ}	Velocity at equilibrium
$V_i^{(des)}(t)$	Desired velocity for subject vehicle i at time t
$V(.)$	Optimal velocity function
V_0	Free speed
V_e	Equilibrium speed spacing function
$V_{\max}$	Maximum speed
$\Delta V_i^{(j)}(t)$	speed difference between the speed of the j^{th} lead vehicle and speed of the subject vehicle i at time instance t, i.e., $\Delta V_i^{(j)}(t) = V_j(t) - V_i(t)$
$X_i(t)$	Position of vehicle i at time instance t
X_0	Stopping distance, including vehicle length (m)
ΔX_c	Safe distance which is considered as constant
$\Delta X_i^{(j)}(t)$	Distance $X_j(t) - X_i(t)$ between lead vehicle j and subject vehicle i at time instance t
$\Delta X_i^{(j,des)}(t)$	Desired following distance (m) for subject vehicle i at time t with respect to lead vehicle j at time instance t
$\Delta X_i^{(jam)}(t)$	Minimum spacing at the standstill situation
$\Delta X_i^{(\min)}$	Minimum space headway for subject vehicle i
$\Delta X_i^{(safe)}(t)$	Safe distance at time instance t
ΔX_{equ}	Space gap at equilibrium
L_i	Length of vehicle i
$LS_i^{(j)}$	Lateral separation between subject vehicle i and lead vehicle j
$LS_{\max}$	Maximum lateral separation

$\Delta a_i^{(j)}(t)$	Acceleration difference between lead vehicle j and subject vehicle i at time t
$d_i^{(comf)}$	Comfortable or desired deceleration
T_i	Reaction time or reaction time delay of subject vehicle i
$T_i^{(des)}$	Desired time headway of vehicle i
T_{Gap}	Time gap
Δt	Update time
$V_i^{(des)}(\Delta X_i^{(j)}(t))$	Desired speed of subject vehicle i for a given space headway $\Delta X_i^{(j)}(t)$ at time t
SV	Subject vehicle
MF	Middle front
LF	Left front
RF	Right front
LS	Left side
RS	Right side
MF1	First vehicle in MF compartment
MF2	Second vehicle in MF compartment
LF1	First vehicle in LF compartment
RF1	First vehicle in RF compartment
LS1	First vehicle in LS compartment
RS1	First vehicle in RS compartment
DM2	Dummy variable for the presence of two or more lead vehicles in the MF compartment
DL1	Dummy variable for the presence of one or more lead vehicles in the LF compartment

DR1	Dummy variable for the presence of one or more lead vehicles in the RF compartment
DLS1	Dummy variable for the presence of one or more lead vehicles in the LS compartment
DRS1	Dummy variable for the presence of one or more lead vehicles in the RS compartment
LS1	First vehicle in left side compartment
RS1	First vehicle in right side compartment
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
u_{qi}^{*}	Utility that the driver of the q^{th} vehicle perceives from choosing a manoeuvring decision i
m_{qi}	Subject vehicle's acceleration (deceleration) value
x_{qi}, z_{qi}	Column vectors of attributes such as spacing and relative speed
ε_{qi}, η_{qi}	Random error terms
$\phi(.)$	Standard normal probability density function
$\Phi(.)$	Standard normal cumulative density function
σ	Scale parameter
MNL	Multinomial logit model
C_{θ}	Copula function with θ as a scalar dependency parameter
CP	Convergence problem
u_{qit}^{*}	Utility that the driver of the q^{th} vehicle perceives from choosing a manoeuvring decision i at time t
m_{qit}	Subject vehicle's acceleration (deceleration) value at time t
x_{qi}, z_{qi}	Column vector of attributes such as spacing and relative speed

ε_{qi}, η_{qi}, ω_{qi}, ψ_{qi}, ξ_{qi}, τ	Random error terms
DW	Watson test statistic
dw_u	Upper critical values
dw_l	Lower critical values
H_0	Null hypothesis
H_A	Alternative hypothesis
L_q	Likelihood expression for individual q
SL_q	Simulated likelihood expression for individual q
SLL_q	Simulated log likelihood expression for individual q
ICLV	Integrated choice and latent variable
EIV	Errors in Variables
IID	Independent and identically distributed
ASC	Alternative specific constant
CRC	Correlated random coefficients
PLN	Power lognormal
APB	Absolute percentage bias
FSSE	Finite sample standard error
ASE	Asymptotic Standard Error
RMSE	Root mean squared error
DGP	Data generating process
ML model	Mixed logit model
ML-ME-PLN model	ML model with multiplicative PLN distributed error terms
ML-RC-PLN model	ML model with PLN distributed uncorrelated random coefficients

ML-CRC-PLN model	ML model with PLN distributed correlated random coefficients
s	Intent
$P_{q(t+\Delta t)}(s)$	Probability that the driver of a subject vehicle q at time $(t+\Delta t)$ intends to make a higher-level decision (or intent) s
x_{qtk}	k^{th} explanatory variable in the set of K traffic environment variables
J	Set of the possible intents
A, C, D	Accelerate, Decelerate, Constant speed
L, S, R	Steer left of longitudinal axis, steer straight along longitudinal axis, steer right of longitudinal axis
AL, AS, AR, DL, DS, DR, CL, CS, CR	Intents
θ_i	Steering angle
$I=\{l_1,l_2,l_3,r_1,r_2,r_3\}$	Set of angles
$A^s_{q(t+\Delta t)i}$	Modelled acceleration for intent s
$A_{q(t+\Delta t)}$	Observed acceleration
α_{s1} , α_{s2}, α_{s3}, α_{s4}	Parameters
λ_i	Dummy variable coded as 1 if a vehicle is present ahead of the SV along the direction i (0 otherwise)
ΔX_{qti}	Space between the SV and the vehicle ahead of SV along direction i
ΔX^{curb}_{qti}	Space between the SV and the curb ahead along the direction i
ΔV_{qti}	Relative velocity between the SV and the vehicle ahead of it along the direction i
V^{des}	Desired longitudinal speed
ΔX^{des}_{qt}	Desired space gap

$\Delta X_{\min}$	Minimum gap drivers prefer to maintain with respect to a vehicle ahead when their vehicle is at rest in jam density situation
L_q	Length of the subject vehicle
AD_{qti}	Change in steering angle from t to $(t+\Delta t)$
$\eta^{s}_{q(t+\Delta t)}$	Random term representing the execution error for intent s,
$G^{s}_{q(t+\Delta t)}$	Choice set
$L_{q(t+\Delta t)}\left(A_{q(t+\Delta t)}, i\right)$	Likelihood function for the observed actions of the driver of vehicle q at time $(t+\Delta t)$
MVN	Multivariate normal

CHAPTER 1 INTRODUCTION

1.1 MOTIVATION

Driver behaviour models are widely used in traffic engineering literature and practice for understanding and describing drivers' manoeuvring decisions in vehicular traffic streams. They also form the building blocks of traffic microsimulation tools, which are used for traffic flow analysis, traffic safety analysis, traffic emission estimation, traffic control studies, etc.

Most driver behaviour models in the literature are built for *homogeneous traffic* conditions typically observed in countries such as Australia, the United States, Germany, and the Netherlands. Such *homogeneous traffic* streams are dominated mostly by passenger cars with similar physical and operational characteristics. Also, vehicles in *homogeneous traffic* streams follow lane discipline and perform limited lateral movements, primarily for lane changing. Various driver behaviour models have been developed to analyse *homogeneous traffic* streams. However, they cannot be directly applied to analyse the *heterogeneous, disorderly* (HD) traffic conditions typically observed in Asian countries such as India, Pakistan, China, Bangladesh, and Indonesia. This is because the characteristics of HD traffic streams tend to be significantly different from those of *homogeneous traffic* streams. HD traffic streams comprise a wide variety of vehicle classes (such as passenger cars, motorcycles, buses, trucks, three-wheeled auto-rickshaws, and non-motorised vehicles) with considerably different physical and operational characteristics. Most of these classes have substantial representation in the traffic streams. Moreover, as opposed to vehicles in *homogeneous traffic* streams, vehicles in HD traffic exhibit weak to no lane discipline, a greater extent of lateral movements, and behaviours such as staggered following and squeezing in the gaps between vehicles (Asaithambi et al., 2016).

Furthermore, regardless of whether the traffic condition is homogeneous or heterogeneous, drivers' decisions are driven by various factors, including human factors, roadway geometry, and traffic environment. Consideration of human factors such as reaction time, desired speed, desired spacing, perception errors, and multi-vehicle anticipation (MVA) are necessary for a realistic representation of driver behaviour (Hamdar, 2012; Treiber and Kesting, 2013; Saifuzzaman and Zheng, 2014). Recent studies have also demonstrated the significance of incorporating human factors in the microscopic driver behaviour modelling framework to comprehensively describe drivers' decision-making process (Saifuzzaman et al., 2015; H. van Lint et al., 2016; Ali et al., 2019; Sharma et al., 2019; Calvert et al., 2020). One such human factor is multi-vehicle anticipation (or MVA, also known as spatial anticipation), which refers to a driver's ability to take several vehicles around their vehicle into account for making their driving decisions. In this context, previous studies argue that considering a single lead vehicle as the only influencer of the subject vehicle's (SV) driving behaviour may not adequately represent SV's driving decisions (Bexelius, 1968; Lenz et al., 1999; Hoogendoorn and Ossen, 2006; Hoogendoorn et al., 2006; Treiber et al., 2006; Peng and Sun, 2010; Zhang, 2014). Yet, the literature on driver behaviour in both HD traffic and *homogeneous traffic* settings has not paid adequate attention to the incorporation of MVA.

Another human factor – drivers' error in perceiving traffic environment variables – has also been recognised as an essential element for improving the realism of driver behaviour models. However, most studies that consider driver's perception error are in the context of a single-leader car-following setting, where the subject vehicle is assumed to follow a single lead vehicle ahead of it and in the context of *homogeneous traffic* conditions. However, it is perhaps much more important to consider drivers' perception errors while analysing driver behaviour in HD traffic streams, for drivers may have to process information from many more sources in HD traffic settings (because of MVA) than in *homogeneous traffic* settings. Even in the context of *homogeneous traffic* conditions, the influence of MVA and drivers' perception errors are still underexplored.

Additionally, lateral movements of vehicles are typically treated in the current literature as a part of lane-changing manoeuvres, which are modelled separately from longitudinal movements. This is again a feature common to models describing *homogeneous traffic* streams with lane discipline. However, in HD traffic streams with weak (or without) lane discipline, a considerable portion of vehicular movements tend to be two-dimensional (2D), with the simultaneity of longitudinal and lateral movements – particularly for motorised two-wheelers and during slow-moving and high vehicle density conditions. Therefore, it is likely that a 2D characterisation of vehicular movements, where both longitudinal and lateral movements are considered simultaneously, might better represent driving behaviour in HD traffic streams.

In view of the above discussion, the overarching goal of this dissertation is to develop driver behaviour models for HD traffic streams on uninterrupted traffic facilities. The proposed driver behaviour models will incorporate the effect of MVA and driver's perception errors and consider the 2D movements of drivers in HD traffic conditions.

The remainder of the chapter is organised in the following manner. Section 1.2 provides a discussion of the gaps in the literature. Section 1.3 outlines the objectives of this dissertation. The overall structure and contributions of this thesis are discussed in Section 1.4. Finally, Section 1.5 identifies the limitations in the scope of the dissertation.

1.2 GAPS IN LITERATURE

Chapter 2 presents a detailed literature review for the reader to understand that a substantial amount of previous research has focused on modelling driver behaviour in HD traffic streams, even if it is not as much as the body of literature on *homogeneous traffic* streams. However, there are still research needs in this area. Specifically, we identify four research gaps, which are discussed next.

1.2.1 Inadequate Consideration of the MVA Effect while Modelling Driver Behaviour in HD Traffic Streams

Our synthesis of the literature revealed that most studies on modelling driver behaviour for HD traffic conditions consider only the effect of a single leader on the subject vehicle. Some studies have considered the influence of multiple vehicles positioned ahead of the subject vehicle only in the longitudinal direction (Jin et al., 2010; Li et al., 2015; Li et al., 2016). However, vehicles that are positioned on either side of the subject vehicle, as well as those positioned ahead in an oblique direction (i.e., vehicles that are not directly ahead), can also influence the driver's driving behaviour. This creates a need to consider the effect of multiple surrounding vehicles (MVA effect) to model driver behaviour in HD traffic conditions.

1.2.2 Limited Efforts to Consider Drivers' Discrete and Continuous Decisions Separately but Model Them Simultaneously

Most driver behaviour modelling studies treat the drivers' decisions – whether to accelerate/decelerate/maintain a constant speed and the decisions on the extent of acceleration or deceleration – all as a single continuum. A single variable or distribution is used to represent all these decisions by treating a positive rate of change of speeds as acceleration and a negative rate of change of speed as deceleration. To the best of the authors' knowledge, only one study (Koutsopoulos and Farah, 2012) separates the decision to accelerate, decelerate or maintain a constant speed (i.e., the discrete decisions, or decisions that may be represented as a discrete variable) and the extent of these decisions such as the extent of acceleration or deceleration (i.e., the continuous decisions, or decisions that may be represented as continuous variables). Although the discrete decisions (of whether to accelerate, decelerate, or maintain a constant speed) and the continuous decisions (of how much to accelerate or decelerate) occur in tandem and without a discernible time lag, it is likely that the cognitive efforts required to decide whether to accelerate, decelerate, or remain in same speed are different from the cognitive efforts needed to determine how much to accelerate or decelerate. Besides, the factors influencing the decision on whether to accelerate or decelerate might be different (or have a different influence on) than the factors influencing the extent of acceleration or deceleration. Hence, there is a need to treat discrete decisions (accelerate, decelerate, or maintain a constant speed) and continuous decisions (how much to accelerate or decelerate) separately and model them as separate entities. At the same time, it is necessary to model all these decisions in a simultaneous manner considering any dependencies among them. This is because several unobserved driver behaviour factors can affect both discrete and continuous decisions of drivers, causing dependency between the mathematical constructs employed to model the decisions. For example, it is possible that common unobserved factors that increase the propensity for taking a discrete decision can also increase/decrease the extent of that decision. Therefore, ignoring the dependency due to unobserved factors, if present, can potentially lead to bias in parameter estimates (because of the error terms being correlated to explanatory variables), distorted interpretations/conclusions, and inferior model fit. These issues motivate the need for joint modelling of the discrete and continuous variables used to represent the above decisions.

To be sure, Koutsopoulos and Farah (2012) also jointly analyse the decisions of acceleration or deceleration along with the extent of acceleration or deceleration; however, their model was developed for *homogeneous traffic* conditions and considered the influence of a single lead vehicle only. To the best of the authors' knowledge, there has not been much research on modelling driver behaviour considering MVA in HD traffic streams, while treating the discrete decisions and the extent of these decisions as separate but simultaneous (joint) decisions with interdependencies.

1.2.3 Inadequate Attention to (and Lack of Methods for) Modelling Drivers' Perception Errors in MVA-Based Driver Behaviour Models

Driver behaviour models typically assume that drivers perceive their traffic environment in the form of spacing and the relative speed with respect to vehicles around them and make their manoeuvring decisions on their perceived values. Since they need to make their manoeuvring decisions in a relatively short time in a continuously evolving environment, they might make errors in perceiving the traffic environment. At the same time, the analyst's measurement of the traffic environment from trajectory datasets used to build driver behaviour models may not be the same as what the drivers perceive. The gap between the perceived and measured values results in errors in the traffic environment variables used in the models. Such *errors in variables (EIV)* introduce additional stochasticity in the models used to describe driver behaviour. It is well recognised in the statistical/econometric literature that ignoring errors in explanatory variables can potentially result in biased parameter estimates and distorted interpretations.

Besides the statistical/econometric issues, consideration of drivers' perception errors has long been recognised as an essential element for improving the realism of driver behaviour models. However, most studies that consider driver's perception error are in the context of a single-leader car-following situation, and most studies on this topic have been in the context of *homogeneous traffic* conditions. On the contrary, the sources of influence (or stimuli) on a subject vehicle are likely to be more than a single vehicle in HD traffic streams. In such scenarios, where drivers process multiple sources of information, perception errors might play an influential role. Besides, given the amount of information to be processed in very short time frames, drivers may deliberately pay greater attention to some elements of the traffic environment and lower attention to some other elements. That is, they might prioritize which elements to perceive carefully and which ones to not pay significant attention to and allocate their cognitive efforts accordingly. Therefore, it is important to consider the driver's perception error in a multi-stimuli-based driver behaviour model in HD traffic streams.

There are methodological challenges in recognising stochasticity or *errors in variables* (EIV) due to drivers' perception errors in driving behaviour models. One methodological challenge is related to the specification of the errors in variables. Most literature on accommodating EIV does so through an additive specification of errors in the variables. Applied to drivers' perception errors, an additive specification implies that the drivers' errors do not depend on the magnitude of the physical quantity they perceive (e.g., spacing, relative speed). As we will discuss and demonstrate in Chapter 5, the additive approach to specifying EIVs does not help in identifying variability due to errors in variables in as many variables as the analyst would prefer, particularly if the driver behaviour models are set up as discrete choice models. Besides, literature in the psychophysics of human perception suggests that errors in human perception of physical quantities depend on the magnitude of the quantity being perceived. In this context, only a handful of studies explore the multiplicative error specification, where the error term specific to a variable is multiplied by the measurement of that variable. Hence, there is a substantial scope to explore the value of multiplicative specifications for drivers' perception errors in driver behaviour models.

1.2.4 Limited Attention to Incorporating Drivers' 2D Movement in HD Traffic Streams while also Considering Drivers' Intentions and the MVA Effect

Much of the driver behaviour modelling literature focuses on modelling drivers' acceleration/deceleration decisions in the longitudinal direction only (i.e., the direction of traffic flow). Lateral movements are considered as discrete events primarily motivated by lane changing or turning needs. This may be because of the focus on modelling driver behaviour in *homogeneous traffic* conditions. As discussed earlier, in HD traffic streams with weak (or without) lane discipline, a considerable portion of vehicular movements tends to be 2D, with the simultaneity of longitudinal and lateral movements. Therefore, it is likely that a 2D characterisation of vehicular movements, where both longitudinal and lateral movements are considered simultaneously, might better represent driving behaviour in HD traffic streams.

Furthermore, the driving actions observed in the vehicle trajectory datasets typically available to the analysts do not reveal the drivers' actual intentions. They only reveal the final driving actions taken on the road, in the form of the acceleration/deceleration extents and the angle of the movement. It is worth noting here that the extent of acceleration/deceleration and 2D movements executed by drivers while travelling on the road are usually a consequence of drivers' intentions to accelerate, decelerate, or maintain a constant speed and to steer to the left of, right of, or straight along the longitudinal axis. However, small acceleration/deceleration values and slight angular deviations from the direction of traffic flow are common even if the drivers intend to maintain a constant speed state and to move in a straight path, respectively. This is due to the difficulty in maintaining a constant speed and avoiding any lateral movement (i.e., angular movement). Nevertheless, the actual intentions of drivers are unobserved or latent to the analyst, and only the outcome of the driver's actions (such as the extent of acceleration and angular deviation from the direction of traffic flow) can be observed in the vehicle trajectory datasets. Therefore, it is useful to consider drivers' intentions while modelling their 2D movements. While some literature exists on modelling drivers' intents (that are latent to the analyst) in *homogeneous traffic* streams, none exists on modelling the driver intents in HD traffic streams while considering 2D movements.

Driving decisions in HD traffic streams involve multifaceted decisions such as: (a) the intention to accelerate, decelerate, or maintain a constant speed, (b) the extent of intended acceleration or deceleration, (c) the intention to steer to the left of, right of, or straight along the traffic flow direction, and (d) the specific angular direction of movement. However, from a cognitive science standpoint, humans are endowed with a limited amount of cognitive resources, such as working memory they need to store and process information for making their decisions (Sweller, 1988). Therefore, drivers might allocate their cognitive resources optimally to quickly make their manoeuvring decisions. Specifically, given the multifaceted decisions drivers need to make in a short timeframe and the complexity of the traffic environment around them, it is plausible that they break down their decision-making into manageable steps for cognitive ease. In this context, there is scope to explore if analysing drivers' intents (which are latent to the analyst) first, followed by the specific actions they take, can help in filling this gap (see Choudhury (2007), Koutsopoulos and Farah (2012), and Choudhury and Islam (2016) for some work in this direction). For instance, it may be that higher-level, strategic decisions – such as the intentions of whether to accelerate, decelerate, or maintain a constant speed and whether to steer to the left of, right of, or keep straight along the longitudinal direction – are made first, followed by lower-level, tactical decisions – such as exactly how much to accelerate or decelerate and which specific angular direction to move along.

Finally, as discussed in detail in our literature review in Chapter 2, while many of the above-discussed aspects may have been considered individually in some studies, we are not aware of driver behaviour models that consider 2D movements and latent intents simultaneously while also incorporating the MVA effect for analysing driver behaviour in HD traffic streams.

1.3 OBJECTIVES OF THE DISSERTATION

The overarching goal of this dissertation is to develop driver behaviour models for HD traffic streams on uninterrupted traffic facilities while considering the following aspects – (1) the MVA behaviour, where drivers' manoeuvring decisions are influenced by multiple vehicles around them, as opposed to a single lead vehicle ahead (2) the treatment of driver behaviour as a combination of different manoeuvring decisions, such as the decision of whether to accelerate, decelerate, or remain in same speed and the decision of the extent of acceleration or deceleration (as opposed to a single, continuous variable representing the driver behaviour),

(3) the incorporation of stochasticity due to drivers' perception errors, and (4) the consideration of 2D movements and driver's intentions (latent to the analyst) simultaneously while also incorporating MVA behaviour. To this end, the methodological objectives of this dissertation are as follows:

> **M1:** To develop a discrete-continuous choice modelling framework for describing car drivers' longitudinal movements in HD traffic conditions. This objective serves the following two purposes: (1) the consideration of the MVA effect while modelling driver behaviour in HD traffic streams, and (2) the consideration of the driver's discrete decisions (i.e., the decision of whether to accelerate, decelerate, or maintain a constant speed, represented as a discrete variable) and continuous decisions (i.e., the extents of acceleration and deceleration, represented as continuous variables) separately but also model them simultaneously.
>
> **M2:** To enhance the above-mentioned discrete-continuous modelling framework to recognise subject vehicle- and driver-specific unobserved factors that influence driver behaviour.
>
> **M3:** To incorporate drivers' perception errors for variables describing the traffic environment in discrete choice-based models of driver behaviour. And to evaluate two different ways of specifying stochasticity due to drivers' errors in their perception of the traffic environment – additive stocasticity and multiplicative stochasticity.
>
> **M4:** To develop an MVA-based latent class framework to simultaneously model 2D movements of motorised two-wheelers. And to develop a latent class framework to incorporate the MVA effect and the driver's intentions (those are latent to the analyst) along two dimensions – (a) the intent to accelerate, decelerate, or maintain a constant speed, and (b) the intent to steer to the left of, right of, or straight along the longitudinal direction.

The substantive objectives of this dissertation are as follows:

> **S1:** To demonstrate the importance of MVA using the model formulated for objective **M1** in two different empirical settings – (1) an HD traffic stream setting using a trajectory dataset from the city of Chennai, India and (2) a *homogeneous traffic* stream setting using a trajectory dataset from the United States of America (USA). Another

objective is to compare and contrast car driver behaviour between HD and *homogeneous traffic* conditions using these two trajectory datasets.

S2: To apply the model formulated for objective **M1** to test whether the factors influencing the decision to accelerate or decelerate are different (or have a different influence on) than the factors influencing the extent of acceleration or deceleration.

S3: To apply the model formulated for objective **M3** to evaluate the importance of incorporating drivers' perception errors in traffic environment variables vis-à-vis allowing unobserved heterogeneity in drivers' response to those variables in driver behaviour models.

S4: To apply the model formulated for objective **M4** to test whether the extent of cognitive efforts invested for making higher-level decisions (drivers' intent to accelerate, decelerate, or maintain a constant speed, and intent to steer to the left of, right of, or straight along the longitudinal direction) are different from those invested for lower-level decisions (decisions of exactly how much to accelerate or decelerate and which specific angular direction to move along).

S5: To evaluate the driver behaviour models developed in this dissertation for their ability to mimic the macroscopic traffic flow properties of uninterrupted traffic streams observed in HD traffic conditions. To achieve this, another objective is to develop a traffic simulator for simulating HD traffic streams using the proposed driver behaviour models.

1.4 THESIS OUTLINE AND CONTRIBUTIONS

The rest of this dissertation is organised into seven additional chapters as follows:

Chapter 2 provides a detailed review of MVA-based driver behaviour models developed for both homogeneous and HD traffic streams.

Chapter 3 addresses the first methodological objective (**M1**) by developing an MVA-based discrete-continuous choice modelling framework to model the longitudinal movement behaviour of car drivers in HD traffic conditions. Specifically, car drivers' longitudinal movement behaviour is analysed by considering influences (or stimuli) from multiple lead vehicles (MVA effect) and infrastructure elements around the drivers' vehicles. To do so, this chapter introduces the concept of an *influence zone*, defined as a hypothetical zone within which the surrounding traffic environment, including vehicles, road boundaries, stationary traffic control devices, etc., influences driver's behaviour. Further, this is perhaps the first time that drivers' discrete and continuous decisions are considered separately but modelled simultaneously while also incorporating the influence of multiple surrounding vehicles when modelling driver behaviour in HD traffic streams. The proposed model, when applied to vehicle trajectory data from a traffic stream in Chennai, India, underscores the importance of considering MVA to describe driving behaviour in HD traffic conditions.

Furthermore, the empirical analysis suggests that the cognitive efforts required to decide whether to accelerate, decelerate, or maintain a constant speed (the discrete decision) may be different from the cognitive efforts needed to determine how much to accelerate or decelerate (extent of the decision). Also, the factors influencing the decision on whether to accelerate or decelerate might be different (or have a different influence on) than the factors influencing the extent of acceleration or deceleration. Thus, the first and second substantive objectives (**S1** and **S2**) of this dissertation are addressed in this chapter.

Chapter 4 contributes to the second methodological objective (**M2**) by enhancing the discrete-continuous choice modelling framework developed in Chapter 3 to incorporate subject vehicle- and driver-specific unobserved factors that influence driver behaviour and then use this enhanced modelling framework to compare and contrast car driver behaviour between HD and *homogeneous traffic* conditions. The empirical analysis demonstrates the value of incorporating MVA in two different empirical settings – one in the context of HD traffic using a vehicle trajectory dataset from Chennai, India and the other in *homogeneous traffic* context using trajectory data from California, USA. The empirical results reveal both similarities and differences in car driver behaviour between *homogeneous traffic* and HD traffic conditions trajectory datasets. Specifically, in both traffic conditions' trajectory data, in addition to vehicles ahead of the subject vehicle in its lane, vehicles ahead in the adjacent lanes influence its driver behaviour. However, side vehicles influence drivers' decision-making only in HD traffic conditions. Also, after an extensive empirical investigation, this chapter recommends the appropriate sizes of influence zones for modelling driver behaviours in HD and *homogeneous traffic* streams. Thus, this chapter also addresses the first substantive objective (**S1**) of this dissertation.

Chapter 5 contributes to the third methodological objective (**M3**) by enhancing the MVA-based driver behaviour model by developing a methodology to incorporate drivers' perception errors in variables describing the traffic environment. In this chapter, an econometric analysis is undertaken to evaluate two different ways of specifying errors in traffic environment variables – (a) the additive specification and (b) the multiplicative specification – in discrete choice models of driver behaviour. It is shown that models with an additive error specification are not econometrically identified to be able to incorporate perception errors in several traffic environment variables, while those with multiplicative error specification are not saddled with theoretical identification problems. The effectiveness of the proposed framework is demonstrated through simulation experiments as well as an empirical application for analysing driver behaviour while considering driver errors in perceiving traffic environment variables. This newly proposed modelling framework also provides insights into the role of perception errors in driver behaviour. For example, first, the empirical application demonstrates that allowing for stochasticity due to perception errors in traffic environment variables is more important than allowing unobserved heterogeneity in drivers' response to those variables. Second, greater variation is 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 directly ahead). This may be because drivers pay greater attention to vehicles directly ahead of their vehicle than those that are not ahead (an indication of how they allocate cognitive efforts). Third, stochasticity due to perception errors for relative longitudinal speeds is found to be greater than that for longitudinal space gaps; perhaps because drivers perceive relative speeds less precisely than space gaps. Fourth, 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. The work in this chapter is geared towards addressing the third substantive objective (**S3**) of this dissertation.

Chapter 6 contributes to the fourth methodological objective (**M4**) by formulating a latent class-based driving behaviour framework that considers drivers' intents for modelling vehicles' 2D movements while considering the MVA effect on these movements in HD traffic situations. Specifically, five extensions are proposed to a typical stimulus-response based driving behaviour framework. First, the subject vehicle's 2D movements are represented as a combination of the angular direction of movement with respect to the longitudinal axis and the magnitude of acceleration or deceleration along the angle. Second, a latent class framework is used to recognize drivers' strategic intents (latent to the analyst) in two dimensions: (a) the intent to accelerate, decelerate, or maintain a constant speed, and (b) the intent to steer to the left of, right of, or straight along the longitudinal axis. It is hypothesized that these higher-level, strategic intents precede lower-level, tactical decisions such as exactly how much to accelerate or decelerate and which specific angular direction to move along. Third, the MVA effect is accommodated to recognize that drivers consider stimuli from multiple vehicles in their vicinity. Fourth, a multi-stimuli model of acceleration is formulated based on the assumption that drivers choose an angle of movement that allows them to move with the highest (lowest) possible longitudinal acceleration (deceleration) if they intend to accelerate (decelerate). Fifth, drivers' execution errors are recognized as the difference between their intended acceleration and executed acceleration. An empirical application of the proposed framework is presented for analysing driver behaviour of motorised two-wheelers using an HD traffic conditions trajectory dataset from Chennai, India.

The empirical results highlight the importance of incorporating MVA and considering driver's intents while modelling 2D movements of motorised two-wheelers in HD traffic conditions. Further, the microscopic traffic environment variables are found to have a stronger influence on drivers' higher-level, strategic intents than on their lower-level, tactical decisions. This may be because drivers invest greater cognitive resources in making their higher-level, decisions (which may be latent to the analyst from typical trajectory datasets) than what they invest in making the lower-level, tactical decisions, such as exactly how much to accelerate or decelerate and which specific angular direction to move along. Thus, the fourth substantive objective (**S4**) of this dissertation is addressed in this chapter.

Chapter 7 contributes to the fifth substantive objective (**S5**) by developing a traffic microsimulation platform for applying the above-developed driver behaviour models of cars and motorised two-wheelers to simulate HD traffic streams. Subsequently, this HD traffic simulator is used to evaluate if the developed driver behaviour models reflect typically observed macroscopic properties of traffic flow.

Chapter 8 concludes the dissertation by recapitulating the findings from and contributions of this research and identifying directions for future research.

1.5 LIMITATIONS IN THE SCOPE OF THE DISSERTATION

This dissertation only models the microscopic movement patterns of cars and motorised two-wheelers in HD traffic streams. Other vehicle types such as auto-rickshaws, buses, and trucks are not in the scope of this dissertation. Typically, in the existing literature, longitudinal driving behaviour is modelled for cars and 2D driving behaviour is modelled for motorised two-wheelers. The same is pursued in this dissertation. Further, the microscopic traffic simulation platform built in this dissertation simulates one-directional HD traffic on a mid-block road section consisting of cars and motorised two-wheelers, which is sufficient to test the proposed driver behaviour modelling framework in this dissertation.

The empirical applications of developed driver behaviour models are used to make informed speculations on how drivers might allocate their cognitive resources for making complex driving decisions. Note that this dissertation does not measure drivers' cognitive efforts directly but compares the strength of influence of traffic environment variables on their driving behaviours to arrive at plausible conclusions on the extent of cognitive efforts invested by drivers in making different manoeuvring decisions.

This dissertation focuses on only two human factors – MVA and drivers' perception errors – while modelling driver behaviour. Other human factors, such as aggressiveness and carefulness in driving, driving behaviours based on anticipating future actions of other vehicles, etc. are not addressed in this dissertation. As importantly, this dissertation reduces driving behaviour to the decisions drivers make in the next time step based on the current traffic environment. By doing so, the time dynamics of driving over several time steps, such as the influence of one's anticipated future actions on current actions, are not considered. Finally, this dissertation does not explicitly consider the interactions of different drivers in a traffic stream sharing and competing for the same space. All these are important limitations in the scope of this dissertation, addressing each of which is an important avenue for future research.

CHAPTER 2 MULTI-VEHICLE ANTICIPATION BASED DRIVER BEHAVIOUR MODELS: A SYNTHESIS OF EXISTING RESEARCH

Abstract

Multi-vehicle anticipation (MVA) refers to drivers' ability to consider stimuli from several vehicles ahead in their manoeuvring decisions such as longitudinal, lateral, and a combination of longitudinal and lateral movements. Studies on MVA identify various advantages of incorporating MVA in driver behaviour models, such as improved behavioural realism, superior numerical soundness, and plausible model outputs. While there are some excellent reviews of single leader driver behaviour models available in the literature, none of them explicitly focuses on models incorporating MVA behaviour despite it being an integral aspect of drivers' manoeuvring decisions. This chapter provides a comprehensive review of MVA based driver behaviour models developed for both homogeneous and heterogeneous, disorderly (HD) traffic streams. Among others, our findings indicate that MVA based driver behaviour models follow a similar pattern of extending the established single-leader car-following models, considering vehicles that are directly ahead (in the same lane), and focussing on a fixed number of vehicles ahead.

Note: The material in this chapter is drawn from the following paper:

Nirmale, S. K., Sharma, A., and Pinjari, A. R. (2022). Multi-vehicle anticipation based driver behaviour models: A synthesis of existing research and future research directions. *(In review with Transportation Letters).*

2.1 INTRODUCTION

Driver behaviour models are used to describe driver behaviours of different types of vehicles and serve as building blocks to describe various traffic scenarios ranging from free flow to congested traffic and city traffic to highway traffic. A few examples of well-known driver behaviour models are Newell's model (Gordon Frank Newell, 1961), Gipps' model (Gipps, 1981), Optimal velocity model (Bando et al., 1995), and Intelligent driver model (Treiber et al., 2000). At large, these models are founded on the stimulus-response framework that was first introduced in the 1950s at the General Motors research laboratories (Chandler et al., 1958; Gazis et al., 1961). According to this framework, a driver's response is proportional to the stimulus from a vehicle ahead. In general, an event or a quantity that evokes a specific response from the driver is called a stimulus. Researchers have considered various stimuli such as speed difference between the subject vehicle and the immediate leader, spacing between the subject vehicle and the immediate leader, and other sources such as gradient, traffic signals, lane closures, lane markings, etc. A majority of available driver behaviour models assume that drivers respond to the stimuli from the immediate lead vehicle only (Brackstone and McDonald, 1999; Saifuzzaman and Zheng, 2014). However, it has been demonstrated theoretically and empirically that drivers anticipate downstream traffic conditions, consider stimuli from multiple vehicles ahead and respond accordingly (Hoogendoorn and Ossen, 2006; Hoogendoorn et al., 2006). Therefore, models that consider stimuli from more than one vehicle to describe driver behaviour offer a higher degree of realism than those that consider stimuli from a single vehicle ahead of the subject vehicle. These driver behaviour models are the focus of this review.

Before moving further, we define some terminologies related to the anticipation by drivers that are commonly and interchangeably used in the literature. Existing literature refers to the following terms in the context of anticipation: multi-anticipative, multi-vehicle anticipation, spatial anticipation, temporal anticipation, and multiple sources of stimuli (or information). The terms 'multi-anticipative', 'multi-vehicle anticipation', and 'spatial anticipation' refer to drivers' ability to consider stimuli from several vehicles ahead in their manoeuvring decisions. The term 'temporal anticipation' refers to drivers' ability to anticipate the traffic situation for the next few seconds and react accordingly. Whereas the term 'multiple sources of stimuli' indicates drivers' ability to consider stimuli from lead vehicles as well as from other sources (mentioned above) in their manoeuvering decisions. This review focuses on models incorporating the multi-vehicle anticipation (MVA) behaviour of drivers.

The first question that arises is why do drivers consider stimuli from several vehicles or anticipate their behaviour? Drivers anticipate the surrounding vehicles' behaviour to adjust their manoeuvring decisions and perform them safely. In *homogeneous traffic* streams, while moving in the longitudinal direction, drivers adjust their acceleration/deceleration and its extent by observing the dynamics of vehicles directly ahead. For instance, by noticing the slowing down of vehicles ahead (second leader or third leader), drivers proactively regulate their acceleration to avoid sudden jerks. When it comes to lane-changing, consideration of stimuli from multiple surrounding vehicles becomes more important because a lane change is a complex and riskier manoeuvre and involves interactions with a greater number of vehicles than car-following. Further, it is well-known that the characteristics of *heterogeneous, disorderly* (HD) traffic streams are substantially different than *homogeneous traffic* streams in terms of traffic composition, lane discipline, and overall driver behaviour (more on this in Section 2.3). One can witness longitudinal, lateral, and a combination of longitudinal and lateral (also known as two-dimensional) movements in HD traffic streams. For a two-dimensional movement to be safe, drivers must consider the dynamics of vehicles that are not only directly ahead but also laterally placed. The anticipation of movements of multiple vehicular becomes more important in HD traffic streams (than in *homogeneous traffic* streams) because other vehicles can cut-in at any time and may cause a safety-critical situation. Hence, it is no exaggeration to say that multi-vehicle anticipation is an integral aspect of drivers' manoeuvring decisions. It makes drivers proactive and thereby better and safe decision-makers (Sharma et al., 2017a).

Studies on MVA enumerate various advantages of incorporating MVA in driver behaviour models such as improved behavioural realism, superior numerical soundness, and plausible parameter estimates and model outputs (Bexelius, 1968; Lenz et al., 1999; Hoogendoorn and Ossen, 2006; Treiber et al., 2006). Despite the importance of considering MVA, models that incorporate MVA are underexplored. Although there are some excellent review studies on driver behaviour models (Brackstone and McDonald, 1999; Toledo, 2007; Moridpour et al., 2010; Saifuzzaman and Zheng, 2014; Asaithambi et al., 2016; Munigety and Mathew, 2016; Das and Maurya, 2018b; Mahapatra et al., 2018; Chakroborty et al., 2019; Azam et al., 2020), none of them comprehensively reviews MVA based driver behaviour models. An in-depth exploration of MVA based models will reveal prevalent theories behind such models, approaches to incorporate MVA behaviour, shortcomings of the available approaches, and a way forward to better mimic the MVA behaviour of drivers.

This chapter, hence, attempts to fill this gap by providing a critical review of MVA based driver behaviour models developed for both homogeneous and HD traffic streams. The remainder of this chapter is organised as follows. Sections 2.2 and 2.3 review models that incorporate MVA by a driver in *homogeneous traffic* conditions and HD traffic conditions, respectively. Section 2.4 provides literature review on driver behaviour models with perception errors. Section 2.5 discusses the main findings from the review of existing literature. Section 2.6 synthesises the relevant research gaps in the current literature to incorporate MVA behaviour. Finally, Section 2.7 concludes this chapter.

2.2 MODELS THAT INCORPORATE MULTI-VEHICLE ANTICIPATION BY DRIVERS IN HOMOGENEOUS TRAFFIC STREAMS

We observe that the majority of MVA based models are extensions of well-known single-leader driver behaviour models such as the Gazis-Herman-Rothery (GHR) car-following model, optimal velocity model (OVM), full velocity difference (FVD) model, and intelligent driver model (IDM). Hence, we review the MVA based driver behaviour models under the classifications of extensions of these well-known single-leader models. The functional forms of these single-leader driver behaviour models are provided in Table 2.1.

2.2.1 Extensions of GHR Car-Following Model

Bexelius (1968) and Gazis et al. (1959) were probably the first to extend the well-known GHR car-following model to consider stimuli from multiple leaders. Eq. (2.1) provides the model formulation:

$$a_i(t+T_i) = \sum_{j=1}^{N} \alpha_j \Delta V_i^{(j)}(t) \tag{2.1}$$

where, $a_i(t+T_i)$ is the acceleration of the subject vehicle i at time $t+T_i$; T_i represents reaction time; N is the number of lead vehicles considered; $\Delta V_i^{(j)}(t)$ represents the speed difference between the speed of the j^{th} lead vehicle $\left(V_j(t)\right)$ and speed of the subject vehicle $\left(V_i(t)\right)$, and α_j is sensitivity coefficient (or weight) to $\Delta V_i^{(j)}(t)$. As can be observed from Eq. (2.1), the MVA is incorporated by considering cumulative weighted stimuli from N vehicles ahead. Setting $N=1$ gives the original GHR car-following model with a single lead vehicle. An advantage of this model is its simplicity. However, the model presented in Eq. (2.1) suffers from the same limitations as the original GHR model. These are: (i) inter-driver heterogeneity is not captured since the model assumes the identical reaction time values for all drivers; and (ii) it assumes that the drivers' manoeuvring actions are only dependent on relative speed. Further, the earlier studies that used this model did not provide empirical evidence for the presence of MVA behaviour (Saifuzzaman and Zheng, 2014).

Table 2.1 Single leader driver behaviour models

Model	**Formulation**
GHR linear car-following model	$a_i(t+T_i)=\alpha_j \Delta V_i^{(1)}(t)$
GHR non-linear car-following model	$a_i(t+T_i)=\alpha\left(V_i(t+T_i)\right)^{\beta}\dfrac{\Delta V_i^{(1)}(t)}{\left(\Delta X_i^{(1)}(t)\right)^{\gamma}}$
Helly's model	$a_i(t+T_i)=\alpha_1 \Delta V_i^{(1)}(t)+\alpha_2\left(\Delta X_i^{(1)}(t)-\Delta X_i^{(des)}(t+T_i)\right)$, where, $\Delta X_i^{(des)}(t+T_i)=\beta_0+\beta_1 V_i(t)+\beta_2 a_i(t)$
Optimal velocity model (OVM)	$a_i(t)=\alpha_j\left[V_{OV}\left(\Delta X_i^{(1)}(t)\right)-V_i(t)\right]$, where, $V_{OV}\left(\Delta X_i^{(1)}(t)\right)=\tanh\left(\Delta X_i^{(1)}(t)-2\right)-\tanh(2)$
Full velocity difference (FVD) model	$.\, a_i(t)=a_0\left[V_{OV}\left(\Delta X_i^{(1)}(t)\right)-V_i(t)\right]+\alpha_1\left(\Delta V_i^{(1)}(t)\right).$ where, $V_{OV}\left(\Delta X_i^{(1)}(t)\right)=V_1+V_2\left(C_1\tanh\left(\Delta X_i^{(1)}(t)-L_i\right)-C_2\right)$
Intelligent Driver Model (IDM)	$a_i(t)=a_i^{\max}\left[1-\left(\dfrac{V_i(t)}{V_i^{(des)}(t)}\right)^{\beta}-\left(\dfrac{\Delta X_i^{(1,\mathrm{des})}(t)}{\Delta X_i^{(1)}(t)}\right)^2\right]$, where, $\Delta X_i^{(1,\mathrm{des})}(t)=C_0+C_1\sqrt{\dfrac{V_i(t)}{V_i^{(des)}(t)}}+V_i(t)T_i^{(des)}-\dfrac{V_i(t)\Delta V_i^{(1)}(t)}{2\sqrt{a_i^{\max}b_i^{conf}}}$

Note: Refer to the list of abbreviations and notations provided on Page xxi for details of the terms used in the above table.

To empirically investigate MVA based models, Hoogendoorn and Ossen (2006), Hoogendoorn et al. (2006) and Zhang (2014) analysed vehicle trajectory data using regression analysis. Hoogendoorn and Ossen (2006) utilised the multivariate linear regression analysis to estimate parameters of Bexelius's multi-anticipatory car-following model (Eq. (2.1)). They empirically demonstrated that drivers not only consider the vehicle directly ahead but also the second leader. Furthermore, they concluded that the degree of driver's reaction to the second leader is dependent on both the types of the following and followed vehicles. Drivers following a truck, for example, had a weaker reaction to the second leader on average than those behind a car, perhaps because drivers cannot easily sight past a truck. Similarly, truck drivers have a significant reaction to the second leader, maybe because truck drivers' elevated vantage point offers a greater sight distance.

Next, Hoogendoorn et al. (2006) proposed the following two modifications to the Bexilus' MVA model.

Bexelius Type 2 model:

$$a_i(t+T_i) = \alpha \min\left\{\Delta V_i^{(1)}(t), \Delta V_i^{(2)}(t), ..., \Delta V_i^{(N)}(t)\right\} \tag{2.2}$$

Bexelius Type 3 model:

$$a_i(t+T_i) = \min\left\{\alpha_1 \Delta V_i^{(1)}(t), \alpha_2 \Delta V_i^{(2)}(t), ..., \alpha_N \Delta V_i^{(N)}(t)\right\} \tag{2.3}$$

These models assume that the driver may only respond to the first, second, or third leader based on the speed difference between their vehicle and the lead vehicles. However, these models do not consider relative spacing as a stimulus and ignore behavioural differences between different manoeuvring decisions (acceleration or deceleration). To fill these gaps, Hoogendoorn et al. (2006) proposed a generalised version of Helly's (1959) model by including a distance-dependent factor in it as provided below:

$$a_i(t+T_i) = \sum_{j=1}^{N_1} \alpha_j \Delta V_i^{(j)}(t) + \sum_{j=1}^{N_2} \beta_j \left(\Delta X_i^{(j)}(t) - \Delta X_i^{(des)}(t+T_i)\right) \tag{2.4}$$

where, $\Delta X_i^{(j)}(t)$ represents the distance between the lead vehicle j and subject vehicle i at time t; $\Delta X_i^{(des)}(t+T_i)$ is desired distance between the subject vehicle i and its j^{th} leader at time $(t+T_i)$. In Eq. (2.4), the additive expression brings in relative speed and spacing as different sources of stimuli. Hoogendoorn et al. (2006) employed the maximum likelihood estimation approach to calibrate the sensitivities (aka, coefficients) of the stimuli from different vehicles, and used a likelihood ratio test to evaluate the benefits of accounting for MVA. Their empirical results suggested that incorporating the MVA effect improved the extent to which the models could explain observed driver behaviour.

Following Hoogendoorn et al. (2006), Zhang (2014) provided more insights on the linear-type MVA driver behaviour model. Specifically, Zhang (2014) proposed a modified generalised linear MVA car-following model (Eq. (2.5)) and addressed two issues – multicollinearity between explanatory variables and the serial correlation of time series data.

$$a_i(t+T_i) = \alpha_0 + \sum_{j=1}^{N_1} \alpha_j \Delta V_i^{(j)}(t) + \sum_{j=1}^{N_2} \beta_j \left(\Delta X_i^{(j)}(t) - \Delta X_i^{(des)}(t+T_i) \right) + \gamma V_i\left(t\right) + \varepsilon_{it} \qquad (2.5)$$

Where, V_i represents speed of the subject vehicle i. This study empirically demonstrated that the traffic congestion level (extremely congested and less congested) affects driver's reactions to different stimuli with respect to different lead vehicles. For example, the study's estimation results depicted that drivers in extremely congested traffic conditions would react to relative speeds with respect to the first, second, and even the third leader.

2.2.2 Extensions of Models Based on Optimal Velocity Model (OVM)

2.2.2.1 Extensions of OVM

The OVM was also widely utilised to mimic the MVA behaviour of drivers. For example, Lenz et al. (1999) extended the OVM to include multiple vehicle responses. Following Bexelius (1968), Lenz et al. (1999) assumed that drivers react to the dynamics of their leading vehicle and an arbitrary number of vehicles ahead with a sensitivity α_j. The mathematical formulation is provided below:

$$a_i(t) = \sum_{j=1}^{N} \alpha_j \left[V_{OV}\left(\frac{\Delta X_i^{(j)}(t)}{j} \right) - V_i(t) \right] \qquad (2.6)$$

Here, $N = 1$ leads to the original optimal velocity model. They borrowed the following functional form from Bando et al. (1998)[1] for $V_{OV}(.)$.

[1] Different studies use different functional forms for the optimal velocity function. To avoid confusion and notation burden, we use a generic symbol $V_{OV}(.)$ for the optimal velocity function throughout the chapter.

$$V_{OV}\left(\Delta X_i^{(1)}(t)\right)=\tanh\left(\Delta X_i^{(1)}(t)-C\right)-\tanh\left(C\right) \text{ where } C \text{ is a constant} \tag{2.7}$$

The study demonstrated increased stability with the MVA based model. Same as in the case of the GHR model, the MVA is incorporated by summing up the weighted stimuli from vehicles ahead. However, a functional form is assumed to compute the sensitivity coefficient α_j.

Notably, a few studies have extended Lenz's model by considering reaction time and desired distance (Hoogendoorn et al., 2006; Hu et al., 2014; Chen et al., 2016). For instance, Hoogendoorn et al. (2006) integrated the reaction time (T_i) as per Eq. (2.8) .

$$a_i(t+T_i)=\sum_{j=1}^{N}\alpha_j\left[V_{OV}\left(\frac{\Delta X_i^{(j)}(t)}{j}\right)-V_i(t)\right] \tag{2.8}$$

Later, Ge et al. (2004) proposed an extended car-following model by incorporating headways of arbitrary numbers of lead vehicles in optimal velocity function itself. The functional form of the model is given below:

$$a_i(t)=V_{OV}\left(\sum_{j=1}^{N}\alpha_j\Delta X_i^{(j)}(t)\right) \tag{2.9}$$

where, α_j is a weighted function of $\Delta X_i^{(j)}(t)$ with the following properties:

1. α_j decreases monotonically as j increases, i.e., $\alpha_1 > \alpha_2$ indicating that the influence of the vehicle ahead of the subject vehicle reduces gradually as the distance between the subject vehicle and lead vehicle increases.
2. α_j takes the following functional form

$$\sum_{j=1}^{N}\alpha_j=1,\quad \alpha_1=1\ \forall j=1,\quad \alpha_j=\begin{cases}\dfrac{6}{7^j} & \text{for } j\neq N\\[2ex] \dfrac{1}{7^{j-1}} & \text{for } j=N.\end{cases} \tag{2.10}$$

The authors performed a simulation-assisted stability analysis and found that MVA further stabilises the traffic flow compared to the single leader model. Additionally, the simulation results from the study confirmed that only the information of three cars ahead of the subject vehicle is enough for cooperative driving. However, empirical evidence is missing from this study.

Wilson et al. (2004) proposed multiple look-ahead models that consider information from multiple leaders as provided below.

Model A:

$$a_i(t) = \alpha_0 \left\{ \sum_{j=1}^{N} \alpha_j V_{OV} \left(\Delta X_i^{(j)}(t) \right) - V_i(t) \right\} \tag{2.11}$$

Model B:

$$a_i(t) = \alpha_0 \left\{ \sum_{j=1}^{N} \alpha_j V_{OV} \left(\frac{\Delta X_i^{(j)}(t)}{j} \right) - V_i(t) \right\} \tag{2.12}$$

In Model A, $\alpha_j > 0$ and $\sum_{j=1}^{N} \alpha_j = 1$ so that uniform flow solutions are the same as in the standard OVM, thus, if the space gap between the subject vehicle i and the lead vehicle j is small, the subject vehicle i will have a lower optimal velocity than the standard OVM. Hence, the subject vehicle i will tend to brake earlier when approaching congested traffic. Model B is taken from Lenz et al. (1999) where the optimal velocity of the subject driver i is given by a weighted sum of the original optimal velocity function, evaluated at a distance to each lead vehicle ahead. The study exhibited that the proposed model did not predict unrealistic values of acceleration. Note that we observe two types of representations when MVA behaviour is considered in the OVM. In Equations (2.11) and (2.12), the sensitivity coefficient is directly multiplied to the optimal velocity function, whereas in Equations (2.6) and (2.8), the sensitivity coefficient is multiplied by the difference between optimal velocity function and velocity of the subject vehicle. Note that formal analysis is missing in the existing literature regarding which formulation is better. We believe that the latter is more behaviourally sound since it considers driver's sensitivity to the stimuli, i.e., the difference between the optimal velocity and the velocity of the subject vehicle rather than the optimal velocity only.

Furthermore, Chen et al. (2012) incorporated driver's reaction time in Wilson et al. (2004)'s model and proposed a multiple look-ahead model with driver reaction delay as below:

$$a_i(t+T_i) = \alpha_0 \left\{ \sum_{j=1}^{N} \alpha_j V_{OV} (\Delta X_i^{(j)}(t)) - V_i(t) \right\} \tag{2.13}$$

Following Ge *et al.* (2004), the following functional form was employed by Chen et al. (2012) for α_j.

$$\sum_{j=1}^{N}\alpha_j = 1, \quad \alpha_j = \begin{cases} \dfrac{N-1}{N^j} & \text{for } j \neq N \\ \dfrac{1}{N^{j-1}} & \text{for } j = N. \end{cases} \tag{2.14}$$

The study utilised the following optimal velocity function from the Bando et al. (1995) study.

$$V_{OV}\left(\Delta X_i^{(J)}(t)\right) = 0.5V_{\max}\left[\tanh\left(\Delta X_i^{(j)}(t) - \Delta X_c\right) + \tanh\left(\Delta X_c\right)\right] \tag{2.15}$$

where, $V_{\max}$ is the maximum velocity and ΔX_c is the safety distance.

Hasebe et al. (2003) also extended the OVM to incorporate MVA behaviour and called this model a forward-looking optimal velocity model (FLOVM). Their formulation is given below:

$$a_i(t) = \alpha_0\left[V_{OV}\left(\Delta X_i^{(1)}(t), V(\Delta X_i^{(2)}(t), ..., V(\Delta X_i^{(J)}(t)\right) - V_i(t)\right] \tag{2.16}$$

The authors employed the optimal velocity function of Bando et al. (1995) and demonstrated that the FLOVM provides greater stability than the original OVM.

2.2.2.2 Extensions of full velocity difference (FVD) model

The FVD model (which is also based on the OVM) was extended by Wang et al. (2006) using the velocity differences of multiple vehicles as provided in Eq. (2.17) and called it a multiple velocity difference (MVD) model.

$$a_i(t) = a_0\left[V_i^{(des)}(\Delta X_i^{(j)}(t)) - V_i(t)\right] + \sum_{j=1}^{m}\alpha_j\left(\Delta V_i^{(j)}(t)\right) \tag{2.17}$$

The MVD model offered more stability and a better suppression of traffic jams than the FVD model. The same strategy of cumulative weighted stimuli is adopted in FVD based extensions too.

Next, by introducing the relative speed of multiple lead vehicles ahead of the subject vehicle in the FVD Model, Li and Liu (2006) proposed a forward-looking relative velocity (FLRV) model as given below:

$$a_i(t+T_i)=V_{OV}\left(\Delta X_i^{(1)}(t),\sum_{j=1}^{N}\alpha_j\Delta V_i^{(j)}(t)\right), where \quad \alpha_j=\left(\frac{1}{5}\right)^j \tag{2.18}$$

The modified optimal velocity function is given below:

$$V_{OV}\left(\Delta X_i^{(1)}(t),\sum_{j=1}^{N}\alpha_j\Delta V_i^{(j)}(t)\right)=\tanh\left(\Delta X_i^{(1)}(t)-4\right)-\tanh\left(4\right)+\lambda\left(\sum_{j=1}^{N}\alpha_j\Delta V_i^{(j)}(t)\right) \tag{2.19}$$

where, λ is a constant that is independent of time, velocity, and position. In addition, this study confirmed that incorporating the relative speed as a stimulus can stabilise the traffic flow, just as in the FLOVM.

Later, Yu et al. (2008) proposed an extended model (Eq. (2.20)) that includes a special case of the original OVM and FVD model.

$$a_i(t)=\alpha_0\left[V_{OV}\left(\Delta X_i^{(j)}(t),\Delta X_i^{(j+1)}(t),...,\Delta X_i^{(j+m-1)}(t)\right)-V_i(t)\right]+\alpha_1\left[\Delta V_i^{(1)}(t)\right] \tag{2.20}$$

The authors employed the following optimal velocity function:

$$V_{OV}\left(\Delta X_i^{(j)}(t),\Delta X_i^{(j+1)}(t),...,\Delta X_i^{(j+m-1)}(t)\right)=0.5V_{\max}\left[\tanh\left(\sum_{j=1}^{m}\alpha_j\Delta X_i^{(j)}(t)-\Delta X_c\right)+\tanh\left(\Delta X_c\right)\right] \tag{2.21}$$

where, ΔX_c is the constant safe distance. This study demonstrated that traffic jams are suppressed more efficiently by considering the headway of more lead vehicles ahead of the subject vehicle and relative speed with respect to the first lead vehicle.

Similarly, Peng and Sun (2010) also modified the FVD model to propose the following MVA model.

$$\begin{aligned}a_i(t)=\alpha_0&\left[V_{OV}(\Delta X_i^{(j)}(t),\Delta X_i^{(j+1)}(t),...,\Delta X_i^{(j+N-1)}(t))-V_i(t)\right]\\&+\beta_0G\left(\Delta V_i^{(j)}(t),\Delta V_i^{(j+1)}(t),...,\Delta V_i^{(j+N-1)}(t)\right)\end{aligned} \tag{2.22}$$

where, $V_{OV}(.)$ is the optimal velocity function and $G(.)$ is assumed as a monotonically increasing function and their formulations are given below:

$$V_{OV}(\Delta X_i^{(j)}(t), \Delta X_i^{(j+1)}(t), ..., \Delta X_i^{(j+N-1)}(t)) = \alpha_1 V^*\left(\Delta X_i^{(j)}(t)\right) + \alpha_2 V^*\left(\Delta X_i^{(j+1)}(t)\right) + ... + \alpha_N V^*\left(\Delta X_i^{(j+N-1)}(t)\right)$$
$$\textit{where,} \quad V^*\left(\Delta X_i^{(j)}(t)\right) = 0.5 V_{\max}\left[\tanh\left(\Delta X_i^{(j)}(t) - \Delta X_c\right) + \tanh\left(\Delta X_c\right)\right] \tag{2.23}$$

$$G\left(\Delta V_i^{(j)}(t), \Delta V_i^{(j+1)}(t), ..., \Delta V_i^{(j+N-1)}(t)\right) = \beta_1 \Delta V_i^{(j)}(t) + \beta_2 \Delta V_i^{(j+1)}(t) + ... + \beta_N \Delta V_i^{(j+N-1)}(t) \tag{2.24}$$

To reflect that the influence of a lead vehicle on the subject vehicle reduces with its distance from the subject vehicle, the authors proposed the following functional form for α_j and β_j

$$\sum_{j=1}^{N} \alpha_j = 1, \quad \alpha_j = \begin{cases} \dfrac{N-1}{N^j} & \text{for } j \neq N, \\ \dfrac{1}{N^{j-1}} & \text{for } j = N, \end{cases} \quad \beta_j = \frac{1}{N^{J-1}} \tag{2.25}$$

where, α_j and β_j are the weighting values representing intensity that driver reacts to its j^{th} lead vehicle. Note that Ge *et al.* (2004) and Chen et al. (2012) implemented a similar functional form for α_j in their study. Numerical simulation results by Peng and Sun (2010) showed that traffic jams are reduced effectively when more lead vehicles are considered.

Further, Jin et al. (2011) proposed an extended car-following model that is based on original OVM and FVD models by adopting multiple velocity differences, as below:

$$a_i(t) = \alpha_0\left(V_{OV}\left(\Delta X_i^{(1)}(t), \Delta X_i^{(2)}(t)\right) - V_i(t)\right) + \alpha_1 \Delta V_i^{(1)}(t) + \alpha_2 \Delta V_i^{(2)}(t) \tag{2.26}$$

The authors used the following generalised optimal velocity function.

$$V_{OV}\left(\Delta X_i^{(1)}(t), \Delta X_i^{(2)}(t)\right) = (1-p)\Delta X_i^{(1)}(t) + p\Delta X_i^{(2)}(t), \text{ where, } 0 \leq p \leq \frac{1}{2} \tag{2.27}$$

Note that $\Delta X_i^{(1)}(t)$ is weighted more as compared to $\Delta X_i^{(2)}(t)$ to reflect that the first leader's influence is greater than the second leader's influence. Further, the study demonstrated that considering the generalised optimal velocity function and velocity differences with respect to multiple vehicles stabilises traffic flow and suppresses traffic jams.

Previously discussed studies use only the space headways and relative velocities with respect to multiple lead vehicles to analyse driver behaviour. However, Li et al. (2011) argued that in addition to space headways and velocity differences, acceleration differences could also influence driver behaviour. Therefore, they proposed the multiple headway, velocity, and acceleration difference model that considered all three types of stimuli. The functional formulation is given below.

$$a_i(t) = \lambda_0 \left[V_{OV} \left(\sum_{j=1}^{N} \alpha_j \Delta X_i^{(j)}(t) \right) - V_i(t) \right] + \lambda_1 \left[\sum_{j=1}^{N} \beta_j \left(\Delta V_i^{(j)}(t) \right) \right] + \lambda_2 \left[\sum_{j=1}^{N} \gamma_j \left(\Delta a_i^{(j)}(t) \right) \right] \quad (2.28)$$

where, $\lambda_0 > 0$, and $\lambda_1, \lambda_2 \in [0,1]$ are different sensitivity coefficients; $\Delta a_i^{(j)}(t)$ represents acceleration difference between the lead vehicle j and subject vehicle i at the time t; α_j, β_j, and γ_j are different weighting coefficients. This study used Eq. (2.15) for $V_{OV}(.)$. Moreover, they assumed α_j as a decreasing function with j, i.e., $\alpha_j > \alpha_{j+1}$ and $\sum_{j=1}^{N} \alpha_j = 1$. The functional form for α_j is the same as that from Peng and Sun (2010). Moreover, Li et al. (2011) illustrated that their model had a greater stable region than that of the FVD model.

2.2.3 Extensions of Intelligent Driver Model (IDM)

To incorporate MVA, Treiber et al. (2006) considered cumulative stimuli between the subject vehicle i and lead vehicle j for N lead vehicles as given below:

$$a_i(t+T_i) = a_i^{\max} \left[1 - \left(\frac{V_i(t)}{V_i^{(des)}(t)} \right)^{\beta} \right] - \sum_{j=1}^{N} a_i^{\max} \left(\frac{\Delta X_i^{(j,\text{des})}(t)}{\sum_{j=1}^{N} \Delta X_i^{(j)}(t)} \right)^2 \quad (2.29)$$

where, $\sum_{j=1}^{N} \Delta X_i^{(j)}(t)$ represents net gaps between the subject vehicle i and lead vehicles j and $\Delta X_i^{(j,\text{des})}(t)$ is represented below:

$$\Delta X_i^{(j,\text{des})}(t) = C_0 + C_1 \sqrt{\frac{V_i(t)}{V_i^{(des)}(t)}} + V_i(t) T_i^{(des)} - \frac{V_i(t) \Delta V_i^{(j)}(t)}{2\sqrt{a_i^{\max} b_i^{conf}}} \quad (2.30)$$

Chen et al. (2010) showed that this formulation leads to an issue at equilibrium flow; i.e., the space gap at equilibrium flow is dependent on the number of lead vehicles leading to different desired headways at the same uniform space gap at equilibrium. To overcome this drawback, Chen et al. (2010) proposed the following IDM-based model:

$$a_i(t+T_i) = a_i^{\max}\left[1-\left(\frac{V_i(t)}{V_i^{(des)}(t)}\right)^{\beta}\right] - \sum_{j=1}^{N} a_i^{\max}\alpha_j\left(\frac{\Delta X_i^{(jam)}(t)+V_i^{equ}(t)T_i}{\Delta X_{equ}}\right)^2 \tag{2.31}$$

where, α_j represents weight coefficients which decrease monotonically as j increases to reflect that the influence of lead vehicles decrease with an increase in the space gaps. They also concluded that the stable region increases after incorporating MVA.

2.2.4 Extension of Collision-Avoidance Models

Eissfeldt & Wagner (2003) extended a collision-avoidance based car-following model based on Gipps (1981), Krauss *et al.* (1996) and Krauss (1998) and aimed at clarifying the role of anticipation in the microscopic traffic model through simulation and analytical calculations. The authors assumed that a driver predicts the worst-case strategy (represented as V_{pred}) that the immediate leader will choose in the next time step. V_{pred} was a function of the leader's desired speed which in turn was a function of the speed of and spacing with respect to the second lead vehicle. Considering that there is a second lead vehicle in front of the first lead vehicle within a distance $\Delta X_1^{(2)}$ driving with speed V_2, then

$$V_{pred} = \max\{V_{des} - \varepsilon a, 0\} \tag{2.32}$$

$$V_{des} = \min\left\{V_1 + a, V_{2,safe}\left(V_2, \Delta X_i^{(2*)}\right), V_{\max}\right\} \tag{2.33}$$

$$V_{safe}\left(V_2, \Delta X_i^{(2*)}\right) = -bT_i + \sqrt{b^2T_i^2 + V_2^2 + 2b\Delta X_i^{(2*)}} \tag{2.34}$$

Then, V_{pred} is used to calculate the safe speed for the subject vehicle i as below:

$$V_{i,safe} = -bT_i + \sqrt{b^2T_i^2 + V_{pre}^2 + 2b\left(\Delta X_i^{(1)} + V_{pred}T_i - \min\left\{V_{pred}T_i, \Delta X_{\text{constant}}\right\}\right)} \tag{2.35}$$

where, a and b represent acceleration and deceleration, respectively. The speed of the first and second lead vehicles are represented as V_1 and V_2, respectively. $\Delta X_i^{(1)}$ is the space gap between the first lead vehicle and subject vehicle i whereas $\Delta X_i^{(2*)}$ is the space gap between the second lead vehicle and the first lead vehicle. The parameter ε represents fluctuations in units of acceleration (a).

2.2.5 Extensions of Piece-Wise Linear Car-Following Model

A piece-wise linear car-following model was extended to incorporate MVA in driver behaviour (Farhi et al., 2012). It was a first-order discrete-time model, where speeds were modelled as a function of spacings. In this study, the minimum form was used rather than an additive form, which is commonly incorporated in the context of MVA based driver behaviour models. Moreover, the spacing to the j^{th} lead vehicle was divided by j to make sensitivities uniform across different lead vehicles. The mathematical formulation is presented below:

$$X_i(t+1) = X_i(t) + \min_{1 \le j \le N} (1+\lambda)^{j-1} \min_{u \in U} \min_{w \in W} \left\{ \alpha_{uw} \left(\frac{\Delta X_i^{(j)}(t)}{j} \right) + \beta_{uw} \right\} \tag{2.36}$$

Here, $t+1$ represents the next discrete time step, α_{uw} and β_{uw}, for $(u,w) \in U \times W$ are parameters to be estimated, and U and W are two finite sets of indices. With this formulation, this study concluded that the minimum form used for considering more than one lead vehicle and with a discounting factor (λ) used to favour the closest leader over distant ones. The proposed model was tested only for single driver trajectory data. The authors reported parameter identification issues when a large number of vehicles (the limit was not reported) were considered. Also, their modelling framework did not account for heterogeneity in driving behaviour.

2.2.6 Lane Changing Models

In lane changing, drivers judge the available gaps in the adjacent lane and opt for the safest gap. Routinely, drivers interact with more than one vehicle in the current and adjacent lanes by considering dynamics (speed, spacing, etc.) of those vehicles in their decision-making. Models describing the lane-changing behaviour by design incorporate multi-vehicle anticipation, and hence, justifiably, not much special effort has been devoted in the literature to consider MVA in lane changing models as is the case in single leader longitudinal movement models. Since some excellent review papers on lane-changing models are already available (Toledo, 2007; Moridpour et al., 2010; Zheng, 2014), we briefly review a few classical studies on lane-changing models instead of a detailed review.

Lane-changing models can be broadly classified into the following categories a) Gipps-type models, b) Utility theory-based models, c) Cellular automata-based models, d) Markov process-based models, e) Hazard-based models, f) Fuzzy-logic based models, and g) game theory-based models. Gipps (1986) was the first to model a driver's lane change behaviour. His model is based on the notion of collision avoidance, treats lane changing as a deterministic process, and ignores inconsistency in driver behaviour over time. Building on Gipp's model's shortcomings, Yang and Koutsopoulos (1996) developed and implemented a probabilistic lane-changing model similar to Gipps' model in the microscopic traffic simulator MITSIM. They defined lane changes as mandatory or discretionary and modelled the lane changing process as four consecutive steps: deciding on a lane change, selecting a target lane, examining an acceptable distance, and performing the lane shift. Another weakness of Gipps model (1986) is the assumption that lane change occurs only when a large enough gap exists in the target lane. However, in heavy or congested traffic, this assumption would be unrealistic. To overcome this limitation, Hidas (2002, 2005) proposed an improved modelling framework to capture the vehicular interaction induced by lane change, which was explicitly classified into three categories based on observations from video-recording vehicular trajectory: free, cooperative, and forced lane changes, to overcome this limitation.

Next, Ahmed et al. (1996) implemented utility theory to model the decision process of lane change. The proposed model structure has four latent levels of decision hierarchy, similar to steps given by Yang and Koutsopoulos (1996). Later, Tomer (2003) provided an integrated modelling framework where the car-following and lane-changing behaviour models were joined together in a single model. The integrated model captured the trade-offs between the utility of being in the correct lane and the speed advantage offered by a faster lane. Furthermore, Kesting et al. (2007) proposed a novel logic for modelling lane changing decisions based on lane change's anticipated advantages and disadvantages. For example, the driver attempts to minimise overall braking induced by lane changes. Recently, game theory has received plenty of attention to model lane change behaviour. Talebpour et al. (2015) and Ali et al. (2019) are a few notable attempts in this regard.

2.3 MODELS THAT INCORPORATE MULTI-VEHICLE ANTICIPATION BY DRIVERS IN HETEROGENEOUS DISORDERLY TRAFFIC STREAMS

Most driver behaviour models assume that vehicles follow lane discipline, follow the centre line of the lane, and consider stimuli solely from vehicles in front since they are developed for *homogeneous traffic* conditions. However, in many developing countries (such as India, Bangladesh, China, etc.), lanes may not be well delineated, or lane discipline may not be effectively maintained, allowing vehicles to occupy any lateral position on the road and encouraging two-dimensional movement for the vehicles to navigate ahead. In addition, vehicles from the side can cut-in to the front of a subject vehicle anytime. Furthermore, HD traffic streams comprise a wide variety of vehicle classes (such as passenger cars, motorbikes, buses, trucks, three-wheeled auto-rickshaws, and non-motorised vehicles) with considerably different physical and operational characteristics. Most of these classes have substantial representation in the traffic streams. In contrast, *homogeneous traffic* streams are dominated mostly by passenger cars having similar physical and operational characteristics. Such distinctive characteristics of HD traffic streams cause differences in driver behaviour between homogeneous and HD traffic streams. The same can be reflected in MVA behaviour. Therefore, in this section, we mainly focus on the studies that have considered the MVA effect while modelling driving behaviour in HD traffic streams. Moreover, we divide this section into two subsections – 1) studies that model only longitudinal movements, and 2) studies that model two-dimensional movements.

2.3.1 Multi-Vehicle Anticipation-Based Driver Behaviour Model for Describing Longitudinal Movements

2.3.1.1 Extensions of FVD model

Jin et al. (2010) proposed a non-lane-based FVD model by taking into consideration the effect of the lateral gap between the subject vehicle and the lead vehicle. It was hypothesised that a driver is affected by the vehicle directly ahead as well as adjacent to it.

$$a_i(t) = \alpha\left[U\left(\Delta X_i^{(1)}(t), \Delta X_i^{(2)}(t)\right) - V_i(t)\right] + \beta\left[G\left(\Delta V_i^{(1)}(t), \Delta V_i^{(2)}(t)\right)\right] \tag{2.37}$$

The following functional forms for $U(.)$ and $G(.)$ are employed in their study:

$$U\left(\Delta X_i^{(1)}(t), \Delta X_i^{(2)}(t)\right) = V_{OV}\left(\left(1 - p_i\right)\Delta X_i^{(1)}(t) + p_i \Delta X_i^{(2)}(t)\right) \tag{2.38}$$

$$G\left(\Delta V_i^{(1)}(t), \Delta V_i^{(2)}(t)\right) = \left(1 - p_i\right)\Delta V_i^{(1)}(t) + p_i \Delta V_i^{(2)}(t) \tag{2.39}$$

$$p_i = \frac{LS_i^{(1)}}{LS_{\max}} \tag{2.40}$$

where, $V_{OV}(.)$ is the optimal velocity function provided in Eq. (2.15); p_i captures the effect of lateral separation distance; $LS_i^{(1)}$ is the lateral separation distance between the centre line passing through the subject vehicle i and the centre line passing through the first lead vehicle (i.e., $j=1$). $LS_{\max}$ is the maximum lateral separation distance beyond which the lead vehicle has no influence on the subject vehicle and set to 3.6 m, i.e., typical lane width. Further, when $LS_i = 0$ (i.e., $p_i = 0$), the proposed model can be simplified as the FVD model. When $LS_i = LS_{\max}$ (i.e., $p_i = 1$), the first lead vehicle is on another lane, and the subject vehicle follows the second lead vehicle. This study demonstrated that incorporating the effect of the lateral gap in the car-following model stabilises traffic flow, suppresses traffic jams, and increases capacity.

The above model assumed that only one side of the subject vehicle had a laterally separated vehicle. However, the subject vehicle may be influenced by vehicles travelling laterally to either side of it. Considering this, Li et al. (2015) developed a two-sided lateral gap FVD model for non-lane discipline traffic as given below:

$$a_i(t) = \alpha\left[U\left(\Delta X_i^{(1)}(t), \Delta X_i^{(2)}(t), \Delta X_i^{(3)}(t)\right) - V_i(t)\right] + \beta\left[G\left(\Delta V_i^{(1)}(t), \Delta V_i^{(2)}(t), \Delta V_i^{(3)}(t)\right)\right] \tag{2.41}$$

It was assumed that the first $(j=1)$, second $(j=2)$, and third lead $(j=3)$ vehicles are travelling on the right front side, left front side, and immediate front of the subject vehicle, respectively. Accordingly, stimuli were calculated for each of the lead vehicles. The following functional forms for $U(.)$ and $G(.)$ were utilised:

$$U\left(\Delta X_i^{(1)}(t), \Delta X_i^{(2)}(t), \Delta X_i^{(3)}(t)\right) = \begin{cases} V\left[(1-2p_i)\Delta X_i^{(1)}(t) + 2p_i \Delta X_i^{(3)}(t)\right] \forall LS_i^{(1)} \in \left(0, 0.5LS_{\max}\right) \\ V\left[(2p_i - 1)\Delta X_i^{(2)}(t) + 2(1-p_i)\Delta X_i^{(3)}(t)\right] \forall LS_i^{(1)} \in \left(0.5LS_{\max}, LS_{\max}\right) \end{cases} \tag{2.42}$$

$$G\left(\Delta V_i^{(1)}(t), \Delta V_i^{(2)}(t), \Delta V_i^{(3)}(t)\right) = \begin{cases} (1-2p_i)\Delta X_i^{(1)}(t) + 2p_i \Delta X_i^{(3)}(t) \forall LS_i^{(1)} \in \left(0, 0.5LS_{\max}\right) \\ (2p_i - 1)\Delta X_i^{(2)}(t) + 2(1-p_i)\Delta X_i^{(3)}(t) \forall LS_i^{(1)} \in \left(0.5LS_{\max}, LS_{\max}\right) \end{cases} \tag{2.43}$$

$$p_i = \frac{LS_i^1}{LS_{\max}} \tag{2.44}$$

Similar to Jin et al. (2010), this study also employed the optimal velocity function as per Eq. (2.15). Stability analysis of the proposed model revealed that the model is more efficient in dissipating perturbation than previous FVD based models.

2.3.1.2 Utility theory-based models

In HD traffic streams, the subject vehicle is surrounded by many vehicles, resulting in multiple stimuli sources from multiple vehicles. Therefore, drivers might react to stimuli from a governing leader. However, only the driver's final actions (such as applied acceleration) are observed from trajectory data, and the governing leader is latent to the analyst. This prompts the need to develop a modelling framework to consider the latent leader while analysing driver behaviour. To do so, a latent leader approach was proposed by Choudhury and Islam (2016) to model acceleration decisions. Particularly, a random utility-based modelling framework was proposed with two components: latent leader component and acceleration component. The former was modelled as a random utility-based discrete choice model, and the probability of the front lead vehicle l (which can be front left, front direct, front right) being a governing leader of the subject vehicle i was expressed as:

$$P\left(l_i\left(t\right)\right) = \frac{\exp\left(\beta^{j} Z_i^{j}\left(t\right)\right)}{\sum_J \exp\left(\beta^{j'} Z_i^{j'}\left(t\right)\right)} \quad j, j' \in J = FL, FD, FR \tag{2.45}$$

where, $Z_i^j\left(t\right)$ represents the explanatory variables associated with the lead vehicle j and β^j is an estimated parameter vector associated with respect to the lead vehicle j. Whereas the acceleration component was modelled using the GM model. Strictly speaking, the latent leader model is not an MVA-based model, as it assumes that single (unknown) leader.

2.3.2 Driver Behaviour Model for Describing Two-Dimensional Movements

Models that can describe simultaneous lateral and longitudinal movements offer more realism in describing HD traffic. Yet, such studies are rare (Mahapatra et al. (2018) and Chakroborty et al. (2019) provide a review of such efforts). In the ensuing paragraphs, a brief discussion of these models is provided, along with some recent studies that were not included in previous reviews.

Chakroborty et al. (2004) developed a comprehensive microscopic model for two-way traffic using a potential field approach. The proposed model had two major components, namely, (1) Steering Response Model (SRM, for predicting the steering angles adoption with time) and (2) Acceleration Response Model (ARM, to predict the rate of acceleration and deceleration over time). Later, a Comprehensive Unidirectional Traffic Simulation Model (CUTSiM) was proposed by Maurya (2007b) that considered the Indian traffic conditions explicitly. The proposed model included a lateral control model component that describes the driver's decision to choose a suitable steering angle based on the hypothesised best path along with its longitudinal control model. However, this model was not extensively calibrated and validated. Furthermore, Kanagaraj and Treiber (2018) proposed a two-dimensional time-continuous model for the mixed traffic flow of motorised and non-motorised vehicles based on the force-field model. The trajectory data from Chennai, India, was utilised to calibrate this model. Recently, Delpiano et al. (2019) developed a two-dimensional microscopic car-following model using the social force approach. They argued that the distance maintained by the driver to avoid collisions in all directions is a critical factor and thus, proposed the multi-directional collision avoidance behaviour model where two-dimensional repulsive force between vehicles was modelled. Specifically, three forces were considered that act on each vehicle: the acceleration force (willingness to accelerate), lane force (tendency to be in the centre in a specific lane), and the repulsive force for collision avoidance. Simulation experiments were also performed to reproduce two-dimensional collision avoidance behaviour. They concluded that the proposed model is a sound starting point for building autonomous vehicles traffic flow models and can improve autonomous driving algorithms.

Next, Mathew et al. (2013) proposed a strip-based approach for the simulation of HD traffic conditions. They developed a simulator named Simulation of Mixed Traffic Mobility using a traditional lane-based simulator SUMO. In the proposed approach, the road was divided into thin strips allowing continuous lateral movement rather than conventional discrete lane changing. They observed that the reduction of strip width increases the throughput indicating improved utilisation of road space.

Furthermore, Lee et al. (2009) proposed an agent-based model to simulate motorcycle behaviour in HD traffic conditions. They developed three models to mimic motorcycle movement patterns, namely, the longitudinal headway model, the oblique and lateral headway model, and the path choice model. The longitudinal headway model described the motorcyclist driving behaviour that they maintain a shorter headway when aligning to the edge of the lead vehicle. The oblique and lateral headway model described the headway distribution of motorcycles when they are following the lead vehicles obliquely. The virtual lane-based movement of motorcycles was modelled as a multinomial logit. Next, Shiomi et al. (2012) proposed a utility theory-based approach to describe the two-dimensional movement of the two-wheelers. This approach captured the characteristics of driver perception of the surrounding traffic situation albeit, it failed to capture the heterogeneity across the drivers.

Following a similar utility theory-based approach, Sarkar et al. (2020) proposed a modelling framework that models driver's two-dimensional movement behaviour. This framework proposed two components: area selection and vehicle movement. In the area selection component, the possible movements of subject vehicles for next time steps were considered, and two-dimensional space ahead of the subject vehicle was divided into the number of realistic cones and treated as choice alternatives for the subject vehicle. A multinomial logit model was used to model the driver's discrete decisions. For the vehicle movement component, a modified intelligent driver model was proposed, which simulates the subject vehicle's next position in the selected direction. Both components of the proposed framework were calibrated separately using real trajectory data set collected in Chennai, India.

2.4 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 stimulus from a lead vehicle ahead of the driver's vehicle. In addition to the vehicle kinematics and traffic environment variables considered in typical car-following models, the literature abounds with studies that highlight the importance of accommodating human factors in these models. The human factors include, for example, drivers' socio-demographic characteristics, physiological factors, personality traits, imperfect driving, driving skills, and driving desires (Hamdar, 2012; Treiber and Kesting, 2013; Saifuzzaman and Zheng, 2014; Sharma, Ali, et al., 2018). In this section, we focus on driver behaviour models that consider errors in drivers' perception of their surrounding traffic environment.

Consideration of drivers' perception errors has long been recognized as an essential element 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 where 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 perception errors for relative positions, speeds, and speed differences. This study concludes that errors in estimation (perception error) are influential on driver behaviour and affects 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. Even in the context of accident analysis, literature (Ding et al., 2019) highlights that accurate perceptions of speeds and distances are critical for avoiding crashes.

In another stream of literature, random utility maximization-based discrete choice models have been used to analyze 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 while modelling driver behaviour in the choice modelling literature to represent variables unobserved to the analyst such as latent plans, latent intent, latent leaders, and reaction time (Choudhury, 2007; Koutsopoulos and Farah, 2012; Choudhury and Islam, 2016); however, not to represent drivers' errors in perceiving the traffic environment variables. Almost all these studies consider a single-leader car-following behaviour. Further, most utility-based choice modelling studies do not consider drivers' perception errors (except, for example, Hamdar et al., 2015).

Many studies discussed above do not demonstrate 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 driver's perception error are in the context of a single-leader car-following setting and in *homogeneous traffic* conditions. Given the lack of lane discipline in *heterogeneous traffic* streams observed in many countries, multiple vehicles around a vehicle, as opposed to a single lead vehicle, might influence its driver's behaviour. In such traffic environments, drivers need to perceive and process multiple sources of stimuli for making their manoeuvering decisions. Therefore, errors are likely to be prevalent in their perception of traffic environment variables such as relative distances and speeds.

2.5 SUMMARY OF FINDINGS

Primarily, MVA based driver behaviour models follow a similar pattern of extending the single-leader car-following model by adding weighted stimuli from vehicles ahead. Interestingly, the influence of only a fixed number (say 3 or 5) of vehicles ahead is considered in these models. Moreover, vehicles far ahead have a low influence on drivers' manoeuvring decisions, and hence, weights on the stimuli decrease as the distance from the subject vehicle increases. Furthermore, the type of vehicle ahead (a car or a truck) also influences the MVA behaviour of drivers and thereby, their manoeuvring decisions. For example, a car driver following a truck is blind to vehicles ahead of the truck (second leader, third leader, etc.) and thus, gives low to no weightage to stimuli from those vehicles. Furthermore, incorporating perception errors in MVA based driver behaviour models improved the behavioural soundness of these models. In addition, all studies have concluded that MVA based models perform better in stabilising the traffic and suppressing traffic oscillations than single lead vehicle-based models. Treiber et al. (2006) revealed that the model with MVA has a higher threshold for traffic stability over the single leader model, i.e., traffic remains stable for higher values of reaction times for the model with MVA, which is consistent with real-world traffic observations.

2.6 RELEVANT RESEARCH GAPS

While there has been a substantial amount of previous research in driver behaviour modelling in HD traffic streams, there are still research needs in this area. Specifically, as discussed in Section 1.2 of Chapter 1, we identify the following research gaps:

1. Inadequate consideration of the MVA effect while modelling driver behaviour in HD traffic streams,
2. Limited efforts to consider driver behaviour as a combination of different manoeuvring decisions, such as the decision of whether to accelerate, decelerate, or remain in same speed (represented as a discrete variable) and the decision of the extent of acceleration or deceleration (represented as continuous variables) – as opposed to using a single, continuous variable to represent all these facets of driver behaviour,
3. Inadequate attention to (and lack of methods for) modelling driver's perception errors in MVA-based driver behaviour models, and

4. Limited attention to incorporating driver's two-dimensional (2D) movement in HD traffic streams while also considering driver's intentions (those are latent to the analyst) and the MVA effect.

Section 1.2 of Chapter 1 already discussed these research gaps in detail.

2.7 CONCLUSIONS

This chapter provides a review of notable attempts to incorporate MVA in driving behaviour models. Multi-vehicle anticipation (MVA) refers to drivers' ability to consider stimuli from several vehicles ahead in their manoeuvring decisions. This chapter reviews models that incorporate MVA by a driver in *homogeneous traffic* conditions and HD traffic conditions. The review of the literature made it possible to identify specific gaps that paved the way for the objectives of this dissertations. As will be seen in the subsequent chapters, several other methodological and empirical contributions have been made in our efforts to achieve the objectives of the dissertation.

CHAPTER 3 A DISCRETE-CONTINUOUS MULTI-VEHICLE ANTICIPATION MODEL OF DRIVING BEHAVIOUR IN HETEROGENEOUS DISORDERLY TRAFFIC CONDITIONS

Abstract

This chapter proposes a multi-vehicle anticipation-based discrete-continuous choice modelling framework for describing driver behaviour in heterogeneous, disorderly (HD) traffic conditions. To incorporate multi-vehicle anticipation, the concept of an influence zone around a vehicle (subject vehicle) is introduced. Vehicles within the influence zone can potentially influence the subject vehicle's driving behaviour. Further, driving decisions are characterized as combination of discrete and continuous components. The discrete component involves the decision to accelerate, decelerate, or maintain constant speed and the continuous component involves the decision of how much to accelerate or decelerate. A copula-based joint modelling framework that allows dependencies between discrete and continuous components is proposed. Such a joint modelling framework recognizes that the discrete and continuous decisions are made simultaneously, and common unobserved factors influence both decisions. Additionally, truncated distributions are employed for the continuous model components to avoid the prediction of unrealistically high acceleration or deceleration values. The parameters of the proposed model are estimated using a trajectory dataset from Chennai, India. The empirical results underscore (a) the importance of considering multi-vehicle anticipation for describing driving behaviour in HD traffic conditions, and (b) the efficacy of the joint discrete-continuous system for modelling driving behaviour. Further, not all traffic environment variables found to influence the discrete decisions were found influential on continuous decisions and vice versa. Moreover, the influence of several variables was found to be stronger on the decision to accelerate or decelerate than on the decision of how much to accelerate or decelerate.

Note: The material in this chapter is drawn from the following paper:

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. https://doi.org/10.1016/j.trc.2021.103144

3.1 INTRODUCTION

Microscopic traffic flow models describe the driving behaviour at the individual vehicle level by mimicking driver's decisions to advance in the longitudinal (vehicle-following) and lateral directions (lane-changing) (refer to Brackstone *et al.* (1999), Toledo (2007), Zheng (2014), Saifuzzaman and Zheng (2014), and Mahapatra *et al.* (2018) for a detailed review of microscopic traffic flow models). These models are used to perform traffic flow analysis, traffic safety analysis, traffic emission estimation and control studies, build traffic simulation tools, etc. (Brackstone and McDonald, 1999; Gartner et al., 2001; Park and Won, 2006; Toledo, 2007; Saifuzzaman and Zheng, 2014; Daiheng Ni, 2015; Chakroborty and Das, 2017). A majority of these models are built for *homogeneous traffic* conditions and cannot be applied to *heterogeneous, disorderly* (HD) traffic where the traffic characteristics are significantly different from the *homogeneous traffic* conditions (Choudhury and Islam, 2016). HD traffic conditions may be characterized by the following three features that are not common to *homogeneous traffic* conditions: (i) a wide variety of vehicles that include motorized two-wheelers, cars, auto-rickshaws, buses, and trucks; (ii) non-lane discipline in vehicular movements; and (iii) a large extent of lateral movements than typically observed in *homogeneous traffic* conditions (Venkatachalam and Gnanavelu, 2009; Kanagaraj et al., 2015; Asaithambi et al., 2016; Choudhury and Islam, 2016; Munigety and Mathew, 2016; Kanagaraj and Treiber, 2018; Mahapatra et al., 2018; Sarkar et al., 2020). The vehicles have broad differences in shapes, sizes, acceleration-deceleration capabilities, and degrees-of-freedom contributing to the heterogeneity in their microscopic movement characteristics. Moreover, drivers often position themselves in between other vehicles in an attempt to make use of the entire available space, thus occupy multiple lanes. Therefore, microscopic vehicular movements in HD traffic conditions depict weak to no lane discipline. In summary, the interactions among vehicles and the resulting manoeuvres they undertake are much more complex in HD traffic conditions (than in *homogeneous traffic* conditions) necessitating the need for developing dedicated microscopic driving behaviour models for such conditions.

Whether the traffic condition is homogenous or heterogeneous, drivers' decisions are driven by various factors including human factors, traffic environment, and roadway infrastructure elements. Human factors are often not given their due importance in microscopic traffic flow models which have made them too simple to explain the complex interaction between the human driver and the traffic environment (Sharma et al., 2017b). In this context, recent studies have demonstrated the significance of incorporating human factors in the microscopic modelling framework to comprehensively describe the driver's decision-making process (Saifuzzaman et al., 2015; Sharma et al., 2017b; J. W. C. van Lint and Calvert, 2018; Ali et al., 2019; Paschalidis et al., 2019; Sharma et al., 2019; Calvert et al., 2020). One such human factor is multi-vehicle anticipation by the driver. Multi-vehicle anticipation (also known as spatial anticipation) refers to a driver's ability to take several vehicles around his/her vehicle into account for making his/her driving decisions. Previous studies argue that considering a single lead vehicle as the only influential vehicle may not adequately represent driving behaviour (Bexelius, 1968; Lenz et al., 1999; Hoogendoorn and Ossen, 2006; Hoogendoorn et al., 2006; Treiber et al., 2006; Peng and Sun, 2010; Zhang, 2014). Moreover, multi-vehicle anticipation stabilizes the traffic flow and reduces the destabilizing effect of reaction time (Treiber et al., 2007). Multi-vehicle anticipation is still underexplored in the context of HD traffic conditions.

The objective of this chapter is to present a microscopic modelling framework that mimics the driving behaviour in the longitudinal direction in HD traffic conditions. Importantly, the presented modelling framework incorporates multi-vehicle anticipation by the driver.

The remainder of this chpater is structured as follows. Section 3.2 presents the research gaps, and positions the contribution of the current chapter. Section 3.3 discusses the proposed modelling framework. Section 3.4 provides an overview of the trajectory dataset, model variables, and their descriptive statistics. Section 3.5 presents and discusses the model estimation results. Finally, Section 3.6 summarises the main findings of this chapter and provides future research directions.

3.2 RESEARCH GAPS AND THE CURRENT STUDY

3.2.1 Research Gaps

Synthesis of the literature provided in Chapter 2 revealed that most studies on modelling driver behaviour for HD traffic conditions consider only the effect of a single leader on the subject vehicle. Many studies have considered the influence of multiple vehicles positioned ahead of the subject vehicle only in the longitudinal direction. However, vehicles that are positioned on either side of the subject vehicle as well as those positioned ahead in oblique direction (i.e., vehicles that are not directly ahead) can also influence of the subject vehicle driver's driving behaviour.

Most driver behaviour modelling studies analyze the longitudinal driving behaviour as a single continuum, where the decision to accelerate/decelerate/maintain same speed and the decisions on the extent of acceleration or deceleration are all treated as a single continuum. A single variable or distribution is used to represent all these decisions by treating a positive rate of change of speeds as acceleration and a negative rate of change of speed as deceleration. To the best of the authors' knowledge, only one study separates the decision to accelerate or decelerate or maintain same speed (i.e., the discrete decisions) and the extent of these decisions (Koutsopoulos and Farah, 2012). Although the discrete decisions and the extent of these decisions occur in tandem and without a discernible time lag, it is likely that the cognitive efforts required to decide whether to accelerate, decelerate, or remain in same speed (the discrete decision) are different from the cognitive efforts needed to determine how much to accelerate or decelerate (extent of the decision). Besides, the factors influencing the decision on whether to accelerate or decelerate might be different (or have a different influence on) than the factors influencing the extent of acceleration or deceleration. Hence, there is a need to treat discrete decisions (accelerate or decelerate or remain in same speed) and continuous decisions (how much to accelerate or decelerate) separately and model them as separate entities. At the same time, it is necessary to model all these decisions in a simultaneous manner considering any dependencies among them. This is because several unobserved driver behaviour factors can affect both drivers' discrete and continuous choices, causing dependency between the mathematical constructs employed to model the decisions. For example, it is possible that common unobserved factors that increase the propensity for taking a discrete decision can also increase/decrease the extent of that discrete decision. Hence, ignoring the dependency due to unobserved factors, if present, can potentially lead to bias in parameter estimates (because of the error terms being correlated to explanatory variables), distorted interpretations/conclusions, and inferior model fit. These issues motivate the need for joint modelling of discrete and continuous decisions.

Koutsopoulos and Farah (2012)'s model joins the decisions of acceleration or deceleration along with the extent of acceleration or deceleration; however, the model was developed for homogenous traffic conditions and considers stimulus from a single lead vehicle only. To the best of the authors' knowledge, there has not been much research on modelling driver behaviour by considering the multi-vehicle anticipation and also treating the discrete decisions and the extent of these decisions as separate but simultaneous (joint) decisions with interdependencies.

3.2.2 Current Study

Motivated by the aforementioned research gaps, this chapter proposes a human factors-based discrete-continuous choice modelling framework for describing driving behaviour in the longitudinal direction under HD traffic conditions. More specifically, a multi-vehicle anticipation-based discrete-continuous choice model of longitudinal driving behavior is presented.

To incorporate multi-vehicle anticipation, first, we introduce the concept of an influence zone surrounding every subject vehicle. Next, we consider stimuli (such as relative speed and space gap) from all vehicles within the influence zone instead of considering stimuli only from the immediate lead vehicle. These multiple surrounding vehicles are from the front (straight ahead, left front, and right front spaces ahead of the subject vehicle) and the left and right sides of the subject vehicle. Further, we consider the lateral gap between the subject vehicle and the pavement edge to capture the effect of roadway elements on driving behaviour. The aspects of driving behaviour analysed at each instance (i.e., time step) are: (1) the decision to accelerate, decelerate, or maintain same speed (a discrete choice component) and (2) the extent of acceleration/deceleration conditional on the vehicle's decision to accelerate/decelerate (a continuous choice component). In addition, the dependency between the discrete and continuous decisions is captured through a copula-based framework that captures correlations due to unobserved factors common to discrete and continuous decisions.

In reality, drivers cannot accelerate or decelerate at any desired rate due to limited physical and operational capabilities of vehicles and due to safety considerations. However, most studies that model acceleration or deceleration behaviour treat the acceleration or deceleration values as unbounded variables that do not prevent the simulation of unrealistically high magnitudes of acceleration or deceleration. In this chapter, we use truncated distributions for acceleration and deceleration values to avoid the prediction of unrealistically high acceleration or deceleration values. While most models prevent unrealistic acceleration (or deceleration) values during simulation, this chapter embeds it within the model structure.

3.3 METHODOLOGY

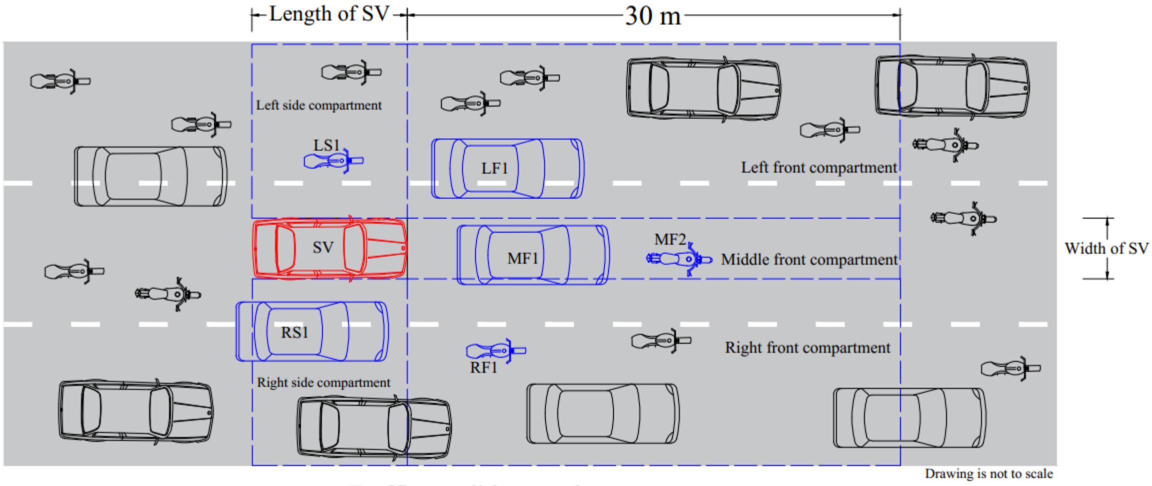
This section discusses the proposed modelling framework under two subsections. In the first subsection, the concept of influence zone around a vehicle is discussed in detail. The second subsection explains the proposed copula-based joint modelling framework and its estimation procedure.

3.3.1 Influence Zone

We define the 'influence zone' of a vehicle as a hypothetical zone within which the surrounding traffic environment including vehicles, road boundary, stationary traffic control devices, etc. influences the driver behaviour. Different sizes and shapes of influence zones may be explored and empirically investigated to ascertain which size and shape provide the best statistical fit to the observed data as well as the most plausible behavioural insights. However, we restrict ourselves to a rectangular influence zone (as shown in Figure 3.1). The reasons behind choosing a rectangular-shaped influence zone are as follows: a) road boundaries for straight sections can be approximated as two edges of a rectangular shaped influence zone and b) its simple geometric properties provide higher computational tractability. For example, one can easily divide the rectangular influence zone into different compartments to examine the effect of different vehicles on the subject vehicle. For the purpose of demonstration in Figure 3.1, the rectangular influence zone is divided into five compartments, namely, middle front (MF), left front (LF), right front (RF), right side (RS), and left side (LS) assuming that the subject vehicle is at the centre of a three-lane roadway. Refer to Figure 3.1 for a diagrammatic representation of these compartments. As the vehicle's longitudinal position shifts to the left or right of the roadway width, the influence zone gets truncated accordingly in that direction. In any case, the compartment in the front of the subject vehicle, extending from the subject vehicle's front bumper to the length of the influence zone is always labeled the MF. When the subject vehicle is at the extreme left edge of the road, there are only three compartments MF, RF, and RS. For the three-lane roadway, the width of the rectangle is approximated to the road pavement width. Particularly, the width of MF is approximated to the width of the subject vehicle, and the width of RF and RS (LF and LS) compartments are set from the right (left) side of the subject vehicle to the right (left) edge of the pavement. Moreover, as shown in Figure 3.1, the back edge of the influence zone follows the back bumper of the subject vehicle.

Next, we also considered the influence of vehicles behind the subject vehicle in this study. However, it is very important to disentangle causality from correlations when ascertaining the influence of a factor (e.g., vehicles behind subject vehicle) on the behaviour of interest (the driving behaviour). Specifically, it is likely that the causal effect of the subject vehicles' actions on that of the vehicles behind them might end up misinterpreted the other way round (i.e., as a causal effect of vehicles behind on the subject vehicle). Statistically speaking, such opposite causality, if not considered, introduces endogeneity in the model parameters and shows as a spurious causal effect. To verify this, we first developed a simple linear model of

SV acceleration as a function of stimuli from all the surrounding vehicles, including those from behind the SV. With such a model, we performed the Hausman test (Hausman, 1978) to examine the endogeneity of the stimuli from the vehicles behind the SV (Wooldridge, 2012). This test indicated that stimuli such as relative speeds with respect to vehicles from behind are endogenous to the SV accelerations. That is, in our empirical data, the effect of the stimuli from vehicles behind on the SV's actions should not be interpreted as a causal effect – it is due to reverse causal effect of the SV's actions on the actions of the vehicles behind. Therefore, we removed the stimuli from vehicles behind from all subsequent model specifications.



Traffic conditions at time *t* sec

Legend

SV : Subject vehicle	RF1 : First vehicle in right front compartment
MF1: First vehicle in middle front compartment	LS1 : First vehicle in left side compartment
MF2: Second vehicle in middle front compartment	RS1 : First vehicle in right side compartment
LF1 : First vehicle in left front compartment	: Influence zone

Figure 3.1 Structure of a rectangular influence zone around a subject vehicle

The next task is to set the length of the rectangular influence zone. In general, the spacing threshold within which vehicles are more likely to be in the vehicle following (or car-following) situation ranges from 30 to 60 m (Herman and Potts, 1959; Subramanian, 1996; Panwai and Dia, 2007; Punzo et al., 2011; Kanagaraj et al., 2015; Sharma, Zheng, et al., 2018; Sarkar et al., 2020). Given the high-density conditions typically observed in HD traffic conditions and to be conservative, we adopt 30 m plus the length of the subject vehicle as the length of the rectangular influence zone in this chapter. The length of RS or LS is equal to the length of the subject vehicle and the length of RF or LF is equal to the length of influence zone minus the length of the subject vehicle. Next, the numbering (i.e., labelling) of a vehicle other than the subject vehicle in the influence zone is based on its longitudinal proximity to the subject vehicle and the compartment to which the vehicle belongs. For example, the label MF1 (MF2) is given to the first closest vehicle (second closest vehicle) in the MF compartment. Each immediate surrounding vehicle within the influence zone is considered in the modelling framework, whether by the side or at the front of the subject vehicle. In addition, the vehicle ahead of the immediate lead vehicle in the middle front zone is also considered as a potentially influencing vehicle. Hence, as shown in Figure 3.1, six potentially influencing vehicles are present. Microscopic traffic environment variables such as space gap and relative speed of the subject vehicle with respect to each of these vehicles (side, right front, left front, and middle front) are considered as potential stimuli for the subject vehicle. These are included as explanatory variables in the modelling framework.

3.3.2 Copula-based Joint Modelling Framework: Formulation and Estimation

Let q be the index representing a driver and let i be the index representing the driver's discrete manoeuvring choice alternatives ($a =$ accelerate, $d =$decelerate, $s =$maintain same speed), at a time instance t. Note that we suppress the subscript t for time instance for ease in notation. In the proposed model, the manoeuvring choice model component takes a random utility-based discrete choice formulation, and the model component for the extent of the manoeuvre takes the form of a truncated normal regression.

As discussed earlier, dependencies may arise between driver's discrete and continuous decisions due to the presence of common unobserved factors influencing both decisions. Such dependencies can be captured through statistical correlations between the stochastic components of the equations used to model drivers' discrete and continuous decisions. In this chapter, we use the copula-based approach to incorporate the dependencies. This approach accommodates non-linear and asymmetric dependencies between different types of marginal distributions (which need not follow the same distribution), while also yielding a closed-form expression for the log-likelihood functions to facilitate parameter estimation. By doing so, the approach overcomes the limitations of typical approaches that can accommodate only linear and symmetric dependencies through linear correlation measures (Pearson's correlation) between marginal distributions of the same type. The copula-based approach has been widely applied in various fields such as econometrics (Cameron et al., 2004; Zimmer and Trivedi, 2006), finance (Embrechts et al., 2002), hydrology (Genest and Favre, 2007) and transportation (Bhat and Eluru, 2009; Spissu et al., 2009). In this chapter, we adopted the copula-based joint discrete–continuous modelling framework proposed by Spissu *et al.* (2009) for simultaneous modelling of the discrete and continuous decisions in the context of driving behaviour. However, the current chapter utilizes truncated distributions for the extent of acceleration and deceleration to recognize physical and safety-motivated limits on how much a driver can accelerate or decelerate. In this section, we first describe the discrete choice and continuous model components separately, followed by the copula-based approach to jointly model the decisions.

Before moving further, it is worth noting that although the driving behaviour is characterized as a set of discrete and continuous decisions, the model does not view driving behaviour as a sequential or hierarchical process where drivers decide on the discrete decisions first and then decide on the continuous decisions. Instead, the copula-based approach is employed to jointly model discrete and continuous decisions, where the model views the decisions as simultaneous in that both the discrete and continuous decision are made jointly.

3.3.2.1 Discrete choice component (for the manoeuvring type choice)

The following equation represents the utility structure of the discrete choice component:

$$u_{qi}^{*} = \beta_i^T x_{qi} + \varepsilon_{qi} \tag{3.1}$$

where, u_{qi}^{*} is the utility that the driver of the q^{th} vehicle perceives from choosing a manoeuvring decision i; x_{qi} is a column vector of attributes (including constant) such as spacing and relative speed between the subject vehicle and all other vehicles in the influence zone, and roadway geometry features that influence the utility for manoeuvring decision i; β_i is the corresponding coefficient vector; and ε_{qi} is the random error term representing the distribution of unobserved factors such as socio-economic characteristics (age, gender, etc.), personality traits (aggressiveness), imperfect driving (perception and estimation errors), etc. influencing u_{qi}^{*}.

A driver of the vehicle q is assumed to choose a manoeuvring decision i if it is associated with the maximum utility among all manoeuvring decisions; that is, if

$$
\begin{aligned}
u_{qi}^* &> \max_{j=a,d,s,j\neq i} u_{qj}^* \\
\beta_i^T x_{qi} + \varepsilon_{qi} &> \max_{j=a,d,s,j\neq i} u_{qj}^* \\
\varepsilon_{qi} - \left\{ \max_{j=a,d,s,j\neq i} u_{qj}^* \right\} &> -\beta_i^T x_{qi} \\
v_{qi} &> -\beta_i^T x_{qi}
\end{aligned}
\tag{3.2}
$$

$$
\text{where, } v_{qi} = \varepsilon_{qi} - \left\{ \max_{j=a,d,s,j\neq i} u_{qj}^* \right\} \tag{3.3}
$$

Eq. (3.2) is equivalent to the multinomial discrete choice model. The distributional assumption of ε_{qi} determines the form of the discrete choice probability expression. As discussed later, this chapter assumes that the ε_{qi} terms are distributed IID type I extreme value across i, which leads to a multinomial logit (MNL) likelihood expression for the discrete choice component.

3.3.2.2 Continuous component (for the extent of acceleration/deceleration)

The model component for the extent of acceleration/deceleration takes the form of the typical linear regression model as below:

$$
m_{qi} = \alpha_i^T z_{qi} + \eta_{qi} \tag{3.4}
$$

where, m_{qi} is the subject vehicle's acceleration (deceleration) value at a time instance t conditional on the decision to accelerate (decelerate). This variable, in turn, is mapped to observed attributes (z_{qi}) influencing the extent of acceleration (deceleration) through the coefficient vectors α_i and a random term η_{qi} that represents the unobserved factors influencing the extent of acceleration (deceleration). To avoid unrealistically high or low acceleration and deceleration values, we use a truncated normal distribution for m_{qi}, which is assumed to arise from a normal distribution with scale $\sigma_{\eta i}$.

3.3.2.3 Copula-based methodology for joint modelling

A copula is a joint distribution function of standard uniform univariate marginals $U[0,1]$ that associates a stochastic multivariate relationship among the univariate marginals. Given two random variables (B_1, B_2) with marginal cumulative distribution functions $(F_1(b_1),\ F_2(b_2))$, their joint bivariate distribution can be represented as follows:

$$
F(b_1, b_2) = C_\theta\left(u_1 = F_1(b_1), u_2 = F_2(b_2)\right) \tag{3.5}
$$

Here, C_θ (copula) is a CDF function which joins $F_1(b_1)$ and $F_2(b_2)$ into a joint distribution.

The individual model components described earlier are brought together, and the likelihood expression of the joint model for a sample of Q vehicles $(q = 1,2,\ldots,Q)$ is represented as below: s

$$L = \prod_{q=1}^{Q}\left[\prod_{i=1}^{I}\left\{P\left(v_{qi} > -\beta_i^T x_{qi}\right)\times P\left(m_{qi}|v_{qi} > -\beta_i^T x_{qi}\right)\right\}^{R_{qi}}\right] \quad \forall i \neq j \tag{3.6}$$

where, $R_{qi} = 1$ if the driver of the q^{th} vehicle chooses alternative i, otherwise $R_{qi} = 0$. I represents total number of alternatives.

3.3.3 Model Estimation

Let $h = \dfrac{m_{qi} - \alpha_i^T z_{qi}}{\sigma_{\eta i}}$. Hence, the conditional likelihood in Eq. (3.6) can be represented as:

$$\begin{aligned} P\left(m_{qi}|v_{qi} > -\beta_i^T x_{qi}\right) &= P\left(\eta_{qi} = h \mid v_{qi} > -\beta_i^T x_{qi}\right) \\ &= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1}\left[P\left(\eta_{qi} = h\right) - P\left(v_{qi} < -\beta_i^T x_{qi}, \eta_{qi} = h\right)\right] \end{aligned} \tag{3.7}$$

A detailed derivation of the above expression is provided in the appendix. As $h = \dfrac{m_{qi} - \alpha_i^T z_{qi}}{\sigma_{\eta i}}$, $\partial h = \dfrac{\partial m_{qi}}{\sigma_{\eta i}} \Rightarrow \sigma_{\eta i}\partial h = \partial m_{qi}$. Therefore, the joint density in Eq. (3.7) can be expressed as:

$$P\left(v_{qi} < -\beta_i^T x_{qi}, \eta_{qi} = h\right) = \frac{1}{\sigma_{\eta i}}\frac{\partial}{\partial h}F_{v_{qi},\eta_{qi}}\left(-\beta_i^T x_{qi}, h\right) \tag{3.8}$$

where $F_{v_{qi},\eta_{qi}}$ represents joint cumulative distribution between random variables v_{qi} and η_{qi}. Now, $P\left(\eta_{qi} = h\right)$ can be expressed as:

$$P\left(\eta_{qi} = h\right) = \frac{1}{\sigma_{\eta i}} f_{\eta i}(h) \tag{3.9}$$

Using Equations (3.8) and (3.9), we can express Eq. (3.7) as below:

$$P\left(m_{qi}|v_{qi} > -\beta_i^T x_{qi}\right) = \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1}\left[\frac{1}{\sigma_{\eta i}} f_{\eta i}(h) - \frac{1}{\sigma_{\eta i}}\frac{\partial}{\partial h}F_{v_{qi},\eta_{qi}}\left(-\beta_i^T x_{qi}, h\right)\right] \tag{3.10}$$

Let the marginal distributions of v_{qi} and η_{qi} be $F_{vi}(.)$ and $F_{\eta i}(.)$, respectively, and the joint distribution of v_{qi} and η_{qi} be $F_{vi,\eta i}(.,.)$. Subsequently, consider $F_{vi,\eta i}(y_1, y_2)$, which can be expressed as a joint cumulative probability distribution of uniform $[0,1]$ marginal variables U_1 and U_2 as below:

$$
\begin{aligned}
F_{vi,\eta i}(y_1, y_2) &= P\left(v_{qi} < y_1, \eta_{qi} < y_2\right) \\
&= P\left(F_{vi}^{-1}(U_1) < y_1, F_{\eta i}^{-1}(U_2) < y_2\right) \\
&= P\left(U_1 < F_{vi}(y_1), U_2 < F_{\eta i}(y_1)\right)
\end{aligned}
\tag{3.11}
$$

Using Sklar's (1973) theorem, a joint K-dimensional distribution of random variables with the continuous marginal CDF functions $F_k(y_k)$ can be generated as below[2]:

$$
\begin{aligned}
F(y_1, y_2, ..., y_k) &= \Pr\left(Y_1 < y_1, Y_1 < y_2, ..., Y < y_k\right) \\
&= \Pr\left(U_1 < F_1(y_1), U_2 < F_2(y_2), ..., U_k < F_k(y_k)\right) \\
&= C_\theta\left(u_1 = F_1(y_1), u_2 = F_2(y_2), ..., u_k = F_k(y_k)\right)
\end{aligned}
\tag{3.12}
$$

Using above theorem, the joint distribution $F_{v_{qi},\eta_{qi}}\left(-\beta_i^T x_{qi}, h\right)$ can be expressed by function $C_\theta(.,.)$ such that:

$$
F_{v_{qi},\eta_{qi}}\left(-\beta_i^T x_{qi}, h\right) = C_\theta\left(u_{q1}^i = F_{v_{qi}}\left(-\beta_i^{'} x_{qi}\right), u_{q2}^i = F_{\eta_{qi}}(h)\right) \tag{3.13}
$$

where $C_\theta(.,.)$ is a copula function and θ is a scalar dependency parameter. This function characterizes the dependency between v_{qi} and η_{qi}. The joint distribution developed in an above-discussed manner is used to derive the likelihood expression. As $u_{q2}^i = F_{\eta_{qi}}(h)$, we can write:

$$
u_{q2}^i = F_{\eta_{qi}}(h) \Rightarrow \partial h = \frac{\partial u_{q2}^i}{f_{\eta i}(h)} \tag{3.14}
$$

---

[2]For better understanding, $\begin{aligned} F_k(y_k) &= \Pr(Y_k < y_k) = \Pr\left(F_k^{-1}(U_k) < y_k\right) = \Pr\left(U_k < F_k(y_k)\right) \\ C_\theta(u_1, u_2, ..., u_k) &= \Pr\left(U_1 < u_1, U_2 < u_2, ..., U_k < u_k\right) \end{aligned}$

Eq. (3.10) can be expressed using Equations (3.13) and (3.14) as below (see Appendix A for details):

$$P\left(\eta_{qi}=h \mid v_{qi}>-\beta_i^T x_{qi}\right)=\left[P\left(v_{qi}>-\beta_i^T x_{qi}\right)\right]^{-1}\left[\frac{1}{\sigma_{\eta i}} f_{\eta i}(h)-\frac{1}{\sigma_{\eta i}} f_{\eta i}(h) \frac{\partial C_\theta\left(u_{q1}^i, u_{q2}^i\right)}{\partial u_2^i}\right] \quad (3.15)$$

Now, substituting Eq. (3.15) into Eq. (3.6), we get

$$L=\prod_{q=1}^{Q}\left[\prod_{i=1}^{I}\left\{\frac{1}{\sigma_{\eta i}} f_{\eta i}\left(\frac{m_{qi}-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)-\frac{1}{\sigma_{\eta i}} f_{\eta i}\left(\frac{m_{qi}-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right) \frac{\partial C_\theta\left(u_{q1}^i, u_{q2}^i\right)}{\partial u_2^i}\right\}^{R_{qi}}\right] \quad \forall i \neq j \qquad (3.16)$$

where,

$$u_{q1}^i=1-\frac{\exp\left(\beta_i^T x_{qi}\right)}{\exp\left(\beta_i^T x_{qi}\right)+\exp\left(\ln \sum_j \exp\left(\beta_j^T x_{qj}\right)\right)} \qquad (3.17)$$

A detailed derivation of the above expression is provided in the Appendix A. A truncated normal regression is used to model the continuous part, u_{q2}^i and can be represented as below:

$$u_{q2}^i=F_{\eta_{qi}}(h)=\frac{\Phi\left(\frac{m_{qi}-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)-\Phi\left(\frac{L_i-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)}{\Phi\left(\frac{U_i-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)-\Phi\left(\frac{L_i-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)} \qquad (3.18)$$

where L_i and U_i are the lower and upper bounds, respectively, on m_{qi}. In the empirical context of this chapter, only two discrete choice alternatives – acceleration and deceleration – have a continuous component. Therefore, the log-likelihood function can be written as follows:

$$L=\prod_{q=1}^{Q}\left[\begin{array}{l}\left\{P\left(m_{qa} \mid \beta_a^T x_{qa}>v_{qa}\right) \times P\left(\beta_a^T x_{qa}>v_{qa}\right)\right\}^{R_{qa}} \times \\ \left\{P\left(m_{qd} \mid \beta_d^T x_{qd}>v_{qd}\right) \times P\left(\beta_d^T x_{qd}>v_{qd}\right)\right\}^{R_{qd}} \times \\ \left\{P\left(\beta_s^T x_{qs}>v_{qs}\right)\right\}^{R_{qs}}\end{array}\right] \qquad (3.19)$$

where, $R_{qa}=1$ if a driver of q^{th} vehicle takes the acceleration decision, 0 otherwise; $R_{qd}=1$ if a driver of q^{th} vehicle takes the deceleration decision, 0 otherwise; and $R_{qs}=1$ if a driver of q^{th} vehicle maintains a constant speed, 0 otherwise.

3.3.3.1 Copula dependence measure

Different copulas offer different levels of ability to capture the dependency between random variables based on the Fréchet–Hoeffding bounds (Bhat and Eluru, 2009). It is useful to construct a scalar metric to measure and compare the dependency implied by different copulas. Traditionally, Pearson's correlation coefficient is used to measure linear dependence between random variables. However, it does not capture asymmetric dependency among random variables. This limitation has led to the use of the concordance measure approach to describe the dependency. Two random variables are said to be concordant (discordant) if large values of one random variable are related to large (small) values of another random variable, and small values of one variable are related to the small (large) values of another random variable. This results in a dependency measure called Kendall's τ, which is defined as the probability of concordance minus the probability of discordance. Kendall's τ satisfies the properties required by Embrechts *et al.* (2002) for assessing the dependency between random variables, such as taking the zero value for independence (i.e., no dependency). Hence, this chapter uses Kendall's τ to characterize and compare the dependency structure. Refer to Embrechts *et al.* (2002) and Bhat and Eluru (2009) for a detailed discussion on different copulas and their implied dependency measures.

3.4 DATA AND VARIABLES CONSIDERED IN THE MODEL

3.4.1 Data

The dataset is a traffic video data of 30 minutes, originally processed by Kanagaraj *et al.* (2015), from an urban arterial stretch in Chennai with HD traffic conditions. This is a straight stretch of 245 meters length and 11.2 meters width (and without any entry or exit locations within or nearby). The processed data include individual trajectories of 3005 vehicles (26.6% cars, 56.4% two-wheelers, and 17% other vehicles such as autorickshaws, trucks, and buses), with each vehicle's trajectory including the spatial position, speed, and acceleration/deceleration values in both the longitudinal and lateral directions at a 0.5 s resolution.

3.4.2 Description of the Explanatory Variables Considered

To analyze a driver's decision to accelerate, decelerate, or maintain constant speed and the extent of acceleration or deceleration at a time instance t s, we considered a variety of traffic environment variables at $t-0.5$ s as sources of stimuli[3]. These variables can be broadly classified into eight groups, as discussed below, and mentioned in Table 3.1.

3.4.2.1 Subject vehicle characteristics

This group includes the subject vehicle's longitudinal speed. As widely recognized in the literature, a vehicle moving at a higher speed is more (less) likely to decelerate (accelerate) and exhibit a higher (lower) magnitude of deceleration (acceleration) than a vehicle moving at a slower speed.

3.4.2.2 Stimuli from the first vehicle in the MF compartment (MF1)

These variables include longitudinal space gap and longitudinal speed difference between the subject vehicle and MF1, MF1 vehicle type (which can modify the stimulus), an interaction between vehicle type and the space gap, and an interaction between vehicle type and the relative speed. A vehicle is more likely to accelerate and have higher acceleration value for larger spacing and higher relative speeds (Chandler et al., 1958; Gazis et al., 1959; Edie, 1961; Gazis et al., 1961; May and Keller, 1967; Ahmed, 1999; Toledo, 2003; Choudhury, 2007; Koutsopoulos and Farah, 2012; L. Li et al., 2013). To account for the indirect influence of the second lead vehicle on the subject vehicle, we considered the acceleration of MF1 at t s as an explanatory variable. The subject vehicle is more likely to accelerate (decelerate) if MF1 has been accelerating (decelerating).

3.4.2.3 Stimuli from the second lead vehicle in the MF compartment (MF2)

This group includes a dummy variable (DM2) for the presence of two or more lead vehicles in the MF compartment, an interaction variable between DM2 and the longitudinal space gap with respect to MF2, and another interaction variable between DM2 and the longitudinal speed difference with respect to MF2. If MF2 is travelling closer to the subject vehicle and at a slower speed than the subject vehicle, the subject vehicle is more likely to decelerate with a higher

[3] To determine the update time, linear regression models of the observed acceleration values as a function of explanatory variables were built considering different update times – ranging from 0.5 s to 2 s. The model with 0.5 s update time provided the best fit to data and the most intuitive interpretations of the parameter estimates. Therefore, an update time of 0.5 s is used in this chapter.

extent of deceleration. Further, the influence of MF2 on the subject vehicle is likely to be less than that of MF1.

3.4.2.4 Stimuli from the first lead vehicle in the LF compartment (LF1)

This category includes the following variables a) a dummy variable (DL1) for the presence of one or more vehicles in the LF compartment, b) an interaction between DL1 and the longitudinal space gap with respect to LF1, c) an interaction between DL1 and the longitudinal speed difference with respect to LF1 and d) an interaction between DL1 and the lateral gap between LF1 and MF1. It is likely that the subject driver chooses to decelerate (and decelerate more) if one or more lead vehicles are present in the LF compartment. In addition, if these lead vehicles are nearby the subject vehicle and are moving at a lower speed, the extent of deceleration is likely to increase. In the same situation, the extent of acceleration is expected to be less if the subject driver accelerates.

3.4.2.5 Stimuli from the first lead vehicle in the RF compartment (RF1)

These variables include a) a dummy variable (DR1) for the presence of one or more lead vehicles in the RF compartment, b) an interaction between DR1 and the longitudinal space gap with respect to RF1, c) an interaction between DR1 and the longitudinal speed difference with respect to RF1, and d) an interaction between DR1 and the lateral gap between the RF1 and MF1. The subject vehicle's driver is likely to decelerate and decelerate more if one or more lead vehicles are present in the RF compartment. Additionally, an increase in the magnitude of deceleration is expected if they are located close to the subject vehicle.

3.4.2.6 Stimuli from the first side vehicle in the LS compartment (LS1)

This group includes a dummy variable (DLS1) to indicate the presence of one or more side vehicles in the LS compartment, an interaction between DLS1 and the lateral space gap between the subject vehicle and LS1, and an interaction between DLS1 and the longitudinal speed difference with respect to LS1. The extent of acceleration is likely to be less if LS1 is laterally closer to the subject vehicle. The speed difference between LS1 and the subject vehicle might also have impact on subject vehicle's decisions.

3.4.2.7 Stimuli from the first side vehicle in the RS compartment (RS1)

The stimuli in this group are a dummy variable for the presence of one or more side vehicles in the RS compartment (DRS1), an interaction between DRS1 and the lateral space gap between the subject vehicle and RS1, and an interaction between DRS1 and longitudinal speed difference with respect to RS1. The extent of deceleration is expected to be high if RS1 is located closer to the subject vehicle while, in the case of acceleration, the magnitude of the acceleration is expected to be low. Same as LS1, the longitudinal speed difference between RS1 and the subject vehicle might influence the subject vehicle's manuevouring decisions.

3.4.2.8 Road geometry characteristics

Road geometry and other elements such as proximity to the road edge, lane demarcations, curves, and surface deformities also influence driver's manoeuvring actions (Maurya, 2007a; Oviedo-Trespalacios et al., 2017). In this chapter, the subject vehicle's proximity to the left edge of the road (i.e., space gap between the left edge of the subject vehicle and the left edge of the road) was included to explore it's influence on driver behaviour. It is anticipated that vehicles closer to the left edge drive slower than those away from the edge. It is worth noting that the proximity of a vehicle to the left edge of the road can potentially be endogenous to its driver behavior. This is because, in the Indian context, vehicles that intend to travel slow tend to keep to the left of the road. Therefore, one must test the endogeneity of this variable before making inferences on the causal influence of proximity a vehicle to the left edge on its driver behaviour.

3.4.3 Descriptive Statistics

The key descriptive statistics of the processed trajectory dataset are given in Table 3.1. As can be observed from the table, in most instances, the subject vehicle in HD traffic condition has at least one vehicle in the LF compartment, possibly because cars typically tend to travel on the right side of the road. Furthermore, the shares of acceleration, deceleration, and maintain same speed states in the Chennai data set are 42.1%, 45.3%, and 12.6%, respectively. This chapter defines acceleration, deceleration, and constant speed (or maintain same speed) states based on following cut-off values of observed acceleration: (a) observed acceleration greater than +0.1 m/s^2 is defined as acceleration state, (b) observed acceleration less than -0.1 m/s^2 is defined as deceleration state, and (c) observed acceleration between -0.1 m/s^2 and +0.1 m/s^2 is defined as constant speed (or maintain same speed) state. In addition, we explored Ozaki's (1993) definition as well, where vehicles accelerating within ±0.05g, where g = 9.8 m/s^2, were considered to be in the constant speed state. Based on this definition, share of acceleration, deceleration, and constant speed states in the Chennai data set are 22.46 %, 23.54 %, and 54.00 %, respectively. It is worth noting here that using Ozaki's definition results in classification of more than half of the data into the constant speed state, which is not likely to be the case in the current empirical context.

3.5 MODEL ESTIMATION RESULTS AND DISCUSSION

The model parameters were estimated using the maximum likelihood approach. A variety of empirical models were estimated, including independent models and different copula-based joint models. The independent model ignores dependencies between the discrete choice and continuous choice components. In the joint model, seven different copulas – Gaussian, FGM, Frank, Gumbel, Clayton, Joe, and AMH copulas – were explored to accommodate dependencies between the discrete and continuous choice components. For modelling the extent of acceleration or deceleration, we explored both unbounded normal and truncated normal distributions to determine which distribution results in and better fit and more behaviourally plausible parameter estimates.

To estimate the model parameters, the manoeuvring decisions of the subject vehicle needs to be determined from the observed data. As mentioned earlier, this chapter classifies various states based on the following cut-off values of acceleration: acceleration greater than +0.1 m/s^2 is defined as acceleration state, acceleration less than -0.1 m/s^2 is defined as deceleration state, and acceleration between -0.1 m/s^2 to +0.1 m/s^2 is defined as constant speed state. We also considered the definition by Ozaki (1993), where the maintain same speed state is defined by acceleration values within ±0.05g. However, some of the estimated model parameters were not behaviourally consistent when we considered Ozaki's (1993) cut-off values. Hence, we do not present and discuss the results based on Ozaki's definition for constant speed.

The literature suggests using +4.00 m/s^2 as upper bound for acceleration and -4.50 m/s^2 as upper bound for deceleration when calibrating the car-following models (HCM, 2000; Saifuzzaman et al., 2015; Sharma et al., 2019). In this chapter, we used +4.15 m/s^2 as upper bound for acceleration and -4.5 m/s^2 as upper bound for deceleration. The reason for not considering +4.00 m/s^2 as the upper bound for acceleration is that we observed acceleration values up to +4.15 m/s^2 in the trajectory data. Therefore, considering only +4.00 m/s^2 as the upper bound would require us to discard a good amount of data with acceleration between +4.00 m/s^2 and +4.15 m/s^2. Therefore, the final truncation thresholds used in this chapter are [-4.5, -0.1) for deceleration, [-0.1, +0.1] for maintain constant speed, and (+0.1, +4.15] for acceleration.

Table 3.1 Descriptive statistics of explanatory variables*

Variables	HD traffic conditions		
	Mean	Std. dev.	Percentage
Dependent variable at t s			
Acceleration decision	--	--	42.10
Deceleration decision	--	--	45.30
Maintain same speed decision	--	--	12.60
Number of vehicles in different compartments of the influence zone			
Subject vehicle has 1 or more vehicles in the middle front (MF) compartment	--	--	100.00
Subject vehicle has 2 or more vehicles in the MF compartment	--	--	39.40
Subject vehicle has 1 or more vehicles in the left front (LF) compartment	--	--	89.50
Subject vehicle has 1 or more vehicles in the right front (RF) compartment	--	--	48.92
Subject vehicle has 1 or more vehicles in the left side (LS) compartment	--	--	47.05
Subject vehicle has 1 or more vehicles in the right side (RS) compartment	--	--	22.20
Type of lead vehicle as a motorcycle in MF compartment			27.30
Subject vehicle (SV) characteristics at t-0.5 s			
Speed in longitudinal direction (m/s)	6.17	1.27	--
Stimuli from MF1 (first vehicle in MF) at t-0.5 s			
Space gap in longitudinal direction (m)	13.94	7.12	--
Relative speed in longitudinal direction (m/s)	0.06	2.70	--
Acceleration at t s (m/s^2)	-0.02	0.85	--
Stimuli from MF2 (second vehicle in MF) at t-0.5 s			
Space gap in longitudinal direction (m)	20.99	5.96	--
Relative speed in longitudinal direction (m/s)	0.18	2.86	--
Stimuli from LF1 (first vehicle in LF) at t-0.5 s			
Space gap in longitudinal direction (m)	8.65	6.94	--
Lateral gap between MF1 and LF1 (m)	2.23	1.32	--
Relative speed in longitudinal direction (m/s)	-0.88	3.12	--
Stimuli from RF1 (first vehicle in RF) at t-0.5 s			
Space gap in longitudinal direction (m)	10.91	8.14	--
Lateral gap between MF1 and RF1 (m)	1.70	1.10	--
Relative speed in longitudinal direction (m/s)	0.67	3.28	--
Stimuli from LS1 (first vehicle in LS) at t-0.5 s			
Space gap in lateral direction (m)	2.79	1.51	--
Relative speed in longitudinal direction (m/s)	-0.86	3.33	--
Stimuli from RS1 (first vehicle in RS) at t-0.5 s			
Space gap in lateral direction (m)	1.65	0.94	--
Relative speed in longitudinal direction (m/s)	0.63	3.31	--
Road geometry characteristics at t-0.5 s			
Space gap between left edge of the subject vehicle and left edge of the road (m)	6.31	1.89	--
Total number of cases	17852		

*Mean and standard deviation of variables with respect to surrounding vehicles are calculated when vehicles are present in the respective compartment

3.5.1 Model Selection

Table 3.2 provides the model fit results for independent and joint models with single leader specifications and multiple leader specifications for unbounded normal and truncated normal distributions of acceleration and deceleration values. This table also presents the statistics for joint models with different copulas. The single leader model specification includes the speed of the subject vehicle, space headway, and relative speed with respect to the immediate lead vehicle, whereas the model with multi-vehicle anticipation considers stimuli from all vehicles that are in the influence zone along with the road geometry characteristics. The different estimated models are not necessarily nested into each other with one as the special case of the other. Therefore, the traditional likelihood ratio test is not applicable to compare the models. Hence, the Bayesian Information Criterion (BIC) metric is employed to compare data fit of the different models. The BIC is formally defined as follows:

$$BIC = K\ln(n) - 2LL(\hat{\beta}) \tag{3.20}$$

where, $LL(\hat{\beta})$ is the log-likelihood value at convergence; K is the number of parameters in the proposed model; and n is number of data points.

A comparison of BIC values suggests that models with multi-vehicle anticipation perform better (lower BIC values) compared to the single leader model specification. Similarly, models with truncated normal distributions of acceleration and deceleration values perform much better compared to those with unbounded acceleration and deceleration values. Since models with truncation on acceleration and deceleration values provided a better fit than those with unbounded normal distributions, we explored these models for further analysis. Overall, the Frank copula-based joint model with truncated normal distributions for acceleration and deceleration values was found to be the best model with the lowest BIC value for the observed data. We also employ an adjusted rho-squared metric, which is used to describe the goodness of fit for statistical models. The adjusted rho-squared is given by:

$$Adj.\,\rho^2 = 1 - \frac{LL(\hat{\beta}) - K}{LL(C) - K_C} \tag{3.21}$$

where, $LL(\hat{\beta})$ is the log-likelihood value at convergence. $LL(C)$ is the log-likelihood value of the constant-only model. K and K_C are the number of parameters in the proposed model and the constant-only model, respectively. As can be observed from Table 3.2, the Frank copula model provides the highest adjusted rho-square value.

Table 3.3 reports the Kendall's τ values used to examine the level of dependency due to unobserved factors influencing the discrete and continuous decisions. As can be noted from the Kendall's τ ranges in the table, not all copulas allow the full possible range of dependency. The Gaussian, FGM, Frank, AMH copulas allow both positive and negative dependencies, whereas the other copulas in the table recognize only positive dependency. Further, the FGM copula allows τ values of only up to $\pm$ 2/9, whereas the AMH copula allows a maximum τ value of 1/3. Such models that allow only positive dependency or those that allow a limited level of dependency provide a poor fit to observed data as evidenced from the BIC values in Table 3.2. Further, models with Joe and AMH copulas were encountered with estimation difficulties because of the limited range of the dependency allowed.

Table 3.2 Estimation results with unbounded normal and truncated normal distributions of acceleration/deceleration values*

Joint model with different copula	Specifications	Copula dependency parameter (t-statistics)		Loglikelihood at convergence	BIC value	Adjusted rho-square
		Acceleration	Deceleration			
Unbounded normal distributions of acceleration/deceleration values						
Independent	SL	-	-	-28972.55	58121.32	0.10
	ML	-	-	-28244.72	57076.83	0.12
Frank copula	SL	-4.36 (-16.71)	-4.518 (-20.02)	-28221.51	56619.24	0.12
	ML	-4.48 (-17.43)	-4.523 (-20.42)	-27506.00	55618.98	0.14
Clayton copula	SL	8.38 (11.34)	9.266 (11.71)	-28043.35	56262.92	0.13
	ML	3.05 (62.89)	2.975 (70.42)	-27724.51	56056.00	0.14
Joe copula	SL	1.00 (10.79)	1.001 (14.84)	-28972.55	58121.32	0.10
	ML	1.00 (13.39)	1.001 (16.72)	-28191.34	57175.66	0.12
AMH copula	SL	-0.97 (-29.94)	-1.000 (-17.64)	-28562.49	57301.20	0.11
	ML	-0.877 (20.8)	-0.961 (-32.33)	-27906.25	56458.63	0.13
Truncated normal distributions of acceleration/deceleration values						
Independent	SL	--	--	-24839.63	49855.48	0.11
	ML	--	--	-24131.08	48898.51	0.13
Gaussian copula	SL	-0.77 (-33.16)	-0.82 (-39.10)	-25064.92	50306.06	0.10
	ML	-0.65 (-29.09)	-0.69 (-55.89)	-24437.07	49422.37	0.12
FGM copula	SL	-0.99 (-15.37)	-1.00 (-20.42)	-24832.49	49841.20	0.11
	ML	-0.77 (-14.50)	-0.84 (-29.54)	-24118.23	48931.54	0.13
Frank copula	SL	-4.65 (11.97)	-5.20 (16.09)	-24726.27	49628.76	0.11
	ML	-4.23 (12.68)	-4.99 (17.73)	-24033.43	48673.84	0.14
Gumbel copula	SL	1.00 (15.13)	2.19 (15.73)	-24867.12	49910.46	0.11
	ML	1.00 (17.58)	1.00 (21.93)	-24120.73	48965.91	0.13
Clayton copula	SL	2.40 (21.32)	2.49 (23.70)	-24900.68	49977.59	0.11
	ML	2.27 (23.53)	2.42 (24.63)	-24329.78	49178.42	0.13

SL-Single leader specification, ML- Multiple leader specification, *The joint copula models with convergence problem are not reported in the table.

The Kendall's τ values from the models that allow both positive and negative dependency (Gaussian, FGM, and Frank copula models) reveal that all the dependency parameters are significantly different from zero, indicating a significant presence of unobserved factors that affect drivers' discrete manoeuvring decisions and the extent of acceleration/deceleration. The most striking result is the negative dependency (or the negative Kendall's τ value) suggesting that the unobserved factors that increase (decrease) the propensity to choose a discrete decision make the driver decrease (increase) the extent of that discrete decision.

Table 3.3 Kendall's dependency measure for different copulas

Copula	Copula parameter (θ)'s range	Kendall's τ	Kendall's τ range	Kendall's τ for acceleration	Kendall's τ for deceleration
Gaussian	[-1,1]	$\tau = \frac{2}{\pi}\arcsin(\theta)$	[-1,1]	-0.450	-0.485
FGM	[-1,1]	$\tau = \frac{2}{9}\theta$	$\left[\frac{-2}{9}, \frac{2}{9}\right]$	-0.171	-0.187
Frank	$[-\infty, \infty]$	$\tau = 1 - \frac{4}{\theta}\left[1 - D_F(\theta)\right]$, $D_{Frank}(\theta) = \frac{1}{\theta}\int_{w=0}^{\theta} \frac{w}{e^w - 1}dw$	[-1,1]	-0.405	-0.456
Gumbel	$[1, \infty)$	$\tau = 1 - \frac{1}{\theta}$	[0,1)	0.001	0.001
Clayton	$(0, \infty)$	$\tau = \frac{\theta}{\theta + 2}$	(0,1)	0.532	0.548
Joe	$[1, \infty)$	$\tau = 1 + \frac{4}{\theta}D_J(\theta)$, $D_{Joe}(\theta) = \int_{w=0}^{1} \frac{\left[\ln\left(1 - w^\theta\right)\right]\left(1 - w^\theta\right)}{w^{\theta - 1}}dw$	[0, 1)	CP	CP
AMH	$[-1, 1]$	$\tau = 1 - \frac{2\left((1-\theta)^2 \log(1-\theta) + \theta\right)}{3\theta^2}$	$\left[\frac{5 - 8\ln 2}{3}, \frac{1}{3}\right]$	CP	CP

CP-Convergence problems encountered in estimating the model

We found similar results when we used Ozaki's (1993) cut-off value of ±0.05g for acceleration magnitude to decide instances of constant speed following. This consistent result across various settings implies that the unobserved factors that make a driver accelerate or decelerate more frequently (less frequently) also make that driver to accelerate or decelerate less (more). The negative dependency can be explained based on unobserved factors such as driver aggressiveness, as follows. As aggressive drivers aim for small gaps with respect to their lead vehicles (Tasca, 2000; Lajunen and Parker, 2001; Abou-Zeid et al., 2011), they tend to accelerate and decelerate more frequently than other vehicles (Ahn and Rakha, 2009). Aggressive drivers do not let the space gap grow much by frequently accelerating to close in on their lead vehicles and then decelerating to maintain the gap. Since such behaviour makes them accelerate and decelerate more frequently and maintain close gaps, the amount of acceleration or deceleration they can attain at any instance tends to be smaller than that of other drivers who tend to wait for longer gaps. Next, one can observe from Table 3.3 that the magnitude of Kendall's τ value is slightly higher for the dependency between the deceleration decision and the extent of deceleration (than that for acceleration related decisions), suggesting that there is a higher level of dependency associated with the deceleration related decisions than that with acceleration related decisions.

3.5.2 Final Model Estimation Results and Discussion

Since the Frank copula model provides the best fit to observed data as well as more behaviourally plausible interpretations, we discuss the estimation results of that model in detail and compare it with the independent model. The estimation results of the independent model and joint model with Frank copula are provided in Table 3.4.

A comparison between the Frank copula and independent models suggests that the parameter estimates from the copula-based joint model have lower t-statistic values when compared to those in the independent model. This is especially so for the continuous components of the model (i.e., for the parameter estimates corresponding to the extent of acceleration and the extent of deceleration). Also, although not reported here, a comparison of various copula-based joint models reveals that the t-statistic values of parameter estimates in the continuous component were found to be smaller for models with stronger copula dependency values. This implies that ignoring negative dependency between the discrete and continuous components may lead to overestimation of the influence of some factors (along with their t-statistic values) on the extent of acceleration or deceleration. This may also lead to a false conclusion that the parameter estimates are more precise. Hence, there will be a tendency not to reject the null hypothesis (parameter does not have an impact on driver's decisions) when it should be rejected. Another point to note is that not all factors that influence the discrete decision (accelerate, decelerate, or maintain constant speed) influence the continuous decisions (how much to accelerate or decelerate) or vice versa. In essence, this discussion highlights the need to model the discrete and continuous components of driver behaviour decisions separately but also reinforces the case for considering the dependency between discrete and continuous decisions. Moreover, the joint model is parsimonious (three less parameters than those in the independent model) and at the same time offers greater explanatory power than the independent model, as indicated by goodness of fit metrics at the end of Table 3.4.

Now, we discuss the influence of different variables on a subject vehicle driver's decisions. In line with expectation, the parameter estimates of the subject vehicle speed variable suggest that vehicles travelling at slower longitudinal speeds are more likely to accelerate (and accelerate more) than vehicles travelling at higher speeds. Similar findings are reported in Siuhi (2009) and Subramanian (1996). Next, the parameter estimates for space headway with respect to MF1 are statistically significant at 1% level of significance for the decisions to accelerate, decelerate, and the extent of acceleration. These results indicate that the subject vehicle is more likely to accelerate and have a higher acceleration value for larger space gaps with respect to the MF1. On the other hand, smaller space gaps make the driver more likely to decelerate and decelerate more. The consideration of dependency between discrete and continuous decisions makes space headway less influential on the extent of deceleration (this parameter had a stronger influence with a higher t-statistic value in the independent model). Next, the parameter estimates of the relative speed are statically significant at 1% significance level in all discrete and continuous decisions with consistent signs as expected. A positive parameter in acceleration decisions (negative parameter in deceleration decisions) suggests that the subject vehicle is more likely to accelerate (decelerate) and accelerate (decelerate) more when MF1 is moving faster than the subject vehicle. The findings are also consistent with those in Koutsopoulos and Farah (2012) in the context of the discrete decision of whether to accelerate or decelerate. For example, the probability of acceleration state increases with an increase in the relative speed between the lead vehicle and the following vehicle.

We have also considered the indirect effect of MF2 (in addition to the direct effect of MF2 discussed later) on the subject vehicle by considering the acceleration of the MF1. It can be seen from Table 3.4 that this variable is statistically significant at 1% significance level in the decision to accelerate, decision to decelerate and the extent of acceleration. The sign of the parameter estimate for this variable is positive in both acceleration decisions and the extent of acceleration, whereas negative in the decision to decelerate. This indicates that the subject vehicle is more likely to accelerate and accelerate more when MF1 is accelerating, and the subject vehicle is more likely to decelerate when MF1 is decelerating.

Furthermore, the effect of the type of immediate lead vehicle was also examined. The parameter for interaction between vehicle type of MF1 and space gap with respect to MF1 has a negative coefficient in the decision to accelerate and the extent of acceleration indicating that the subject vehicle is less likely to accelerate and would accelerate less if MF1 is motorcycle instead of other vehicle types. Similar conclusions can be made for the interaction variable between vehicle type and the relative speed with respect to MF1. Choudhury and Islam (2016) and Ravishankar and Mathew (2011) also report similar influences of the type of lead vehicles on drivers' extent of acceleration.

The estimation results reveal that six out of twelve parameter estimates on the variables related to MF2 are statistically significant. Importantly, no parameter estimate is statistically significant in the extent of deceleration, implying that the presence of a second lead vehicle does not impact the extent of deceleration. Similar observations can be made in the context of variables related to LF1, RF1, LS1 and RS1. Specifically, the extent of deceleration is not influenced as much by the variables corresponding to all these vehicles than by the variables corresponding to MF1. That is, drivers' decisions on the extent of acceleration are governed by multiple vehicles ahead, whereas their decisions on the extent of deceleration are likely to be affected much more by the immediate lead vehicle than that of other vehicles in the influence zone. Several studies (Bexelius, 1968; Lenz et al., 1999; Hoogendoorn and Ossen, 2006) also confirm that drivers respond to the multiple vehicles directly ahead, but we are not aware of studies that report such nuanced findings for the influence of different lead vehicles on driving behaviours. Furthermore, the independent model estimated in this chapter spuriously shows the influence of variables related MF2 on the extent of subject vehicle's deceleration. The joint model, on the other hand, shows no influence of variables related to MF2 on the extent of deceleration. These results highlight the need for considering dependency between discrete and continuous decisions of whether to accelerate/decelerate and how much to do so.

We now turn to the effect of the vehicles in the left front (LF1), right front (RF1), left side (LS1), and right side (RS1) on the subject vehicle's driving behaviour. For each of these vehicles, the parameter estimates reported in Table 3.4 are in line with our expectations. For example, the parameter estimates for the relative speed with respect to LF1 and those with respect to RF1 take a positive sign in the decisions to accelerate and the extent of acceleration.

This implies that the likelihood of acceleration increases with an increase in relative speed between the influencing vehicles (LF1 and RF1) and the subject vehicle. Furthermore, the parameter estimates related to the lateral gap with respect to LS1 indicate that larger (smaller) space gaps are associated with a higher (lower) likelihood of acceleration as well as higher (lower) extent of acceleration. Additionally, the parameter estimates of relative speed with respect to LS1 suggest that drivers are less likely to decelerate and the extent of acceleration increases with an increase in the relative speed between LS1 and the subject vehicle. Also, the parameter estimate for relative speed with respect to RS1 indicates that an increase in the relative speed with respect to RS1 decreases the extent of deceleration. Such findings indicate that drivers' anticipation of traffic situation and their response depends on vehicles on the side in addition to vehicles ahead. Hence, one can conclude that vehicles, which are in the left front, right front, left side and right side, are also likely to affect a driver's decisions.

The gap between the subject vehicle and the left edge of the road was also found to have a significant influence on driving decisions. A smaller gap between the vehicle and the road edge makes car drivers tend to decelerate more. This is intuitively aligned with our expectations since in HD traffic scenarios, for example, in the Indian driving context, low speed traffic opts the left side of the road whereas high speed traffic is on the right side. As discussed earlier, this variable can potentially be endogenous to drivers' decisions to acceleration and deceleration. To correct for such endogeneity, we used the control function method. Specifically, we first regressed the vehicles' distance to the road's left and then used the residual from this regression as another variable in the copula-based joint model. The coefficient of the residual turned out to be statistically insignificant in both discrete and continuous decision equations of the joint model. Therefore, we dropped the endogeneity correction from the final specification of the proposed model.

It is worth noting here that one cannot include another variable titled gap from the right edge of the road, as such a model cannot be identified since the gap to the left edge and that to the right edge (along with vehicle width) would add up to a constant (road width). Importantly, the modelling framework is general enough to allow for the inclusion of roadway geometry elements such as spacing from the curb, presence and spacing of any fixed objects on the road, and irregularities in the geometry as more empirical data becomes available from such varied conditions.

To further demonstrate the importance of considering multi-vehicle anticipation, additional empirical models were estimated by cumulatively including stimuli from one vehicle at a time until all stimuli from all surrounding vehicles in the influence zone were included in the model. Log-likelihood ratio tests were performed to compare these different models (see Table 3.5). As can be observed from the test results in Table 3.5, considering the influence of multiple vehicles improves the model fit. In fact, the model fit improved due to the consideration of each (and every) additional influencing vehicle. This observation underscores the importance of considering multiple-vehicle anticipation in driver behaviour modelling. This finding also validates the existence of the influence zone for each vehicle.

The empirical model presented here extends the current understanding of driving behaviour in HD traffic conditions. The results underscore the importance of considering multi-vehicle anticipation to model driving behaviour, where multiple vehicles within the influence zone of the subject vehicle influences its driver's decisions. Doing so can help improve the realism of driver behaviour models for HD traffic traffic streams. Further, the literature identifies other advantages of considering multi-vehicle anticipation such as an increase in the stability of traffic flow, compensation for the adverse effect induced by reaction delays, and better representation of fundamental traffic flow diagrams (Lenz et al., 1999; Treiber et al., 2006; Wang et al., 2006; Ossen, 2008; Peng and Sun, 2010).

Moreover, as driver behaviour models are vital for traffic flow analysis, traffic safety analysis, and traffic emissions estimation, employing the proposed model can assist in getting more realistic results from such analyses. Given the potential gains from considering multi-vehicle anticipation, the available driver assistance systems technology for HD traffic conditions can potentially benefit from including multi-vehicle anticipation-based prediction algorithms. In addition to HD traffic conditions, the proposed model can be applied in homogenous traffic conditions as well, since (a) multi-vehicle anticipation has been found in homogenous conditions too (Bexelius, 1968; Lenz et al., 1999; Hoogendoorn and Ossen, 2006; Treiber et al., 2006; Peng and Sun, 2010), (b) longitudinal movements dominate in these conditions, and (c) the modelling framework is generic and transferrable. Furthermore, using the proposed model to simulate human-driven vehicles in a mixed traffic of human-driven and autonomous vehicles will likely improve the realism of the simulated mixed traffic environment. Such a simulation environment can potentially offer a more robust investigation of autonomous vehicles' trajectory planning algorithms.

Table 3.4 Estimation results of the independent model and joint model with Frank copula

Explanatory variables	**Independent model**				**Joint model with Frank copula**			
	MNL#		**Truncated regression***		**MNL#**		**Truncated regression***	
	Acceleration	**Deceleration**	**Acceleration**	**Deceleration**	**Acceleration**	**Deceleration**	**Acceleration**	**Deceleration**
Constant	1.371 (7.88)	-0.581 (-3.19)	0.362 (2.26)	-2.395 (-12.59)	1.355 (7.88)	-0.550 (-3.07)	0.681 (4.68)	-0.798 (-5.50)
Subject vehicle (SV) characteristics								
Speed in longitudinal direction (m/s)	-0.038 (-2.67)	0.176 (13.45)	-0.133 (-11.39)	0.185 (18.16)	-0.034 (-2.46)	0.174 (13.52)	-0.047 (-4.50)	0.108 (13.24)
Stimuli from MF1 (first vehicle in MF)								
Space gap in longitudinal direction (m)	0.020 (4.93)	-0.015 (-4.06)	0.027 (8.54)	-0.012 (-4.46)	0.018 (4.52)	-0.014 (-3.81)	0.015 (5.42)	-0.003 (-1.13)
Relative speed in longitudinal direction (m/s)	0.193 (11.62)	-0.151 (-8.97)	0.191 (18.00)	-0.122 (-13.66)	0.178 (10.79)	-0.135 (-8.19)	0.097 (10.25)	-0.020 (-2.36)
Acceleration at t s (m/s^2)	0.311 (7.80)	-0.122 (-3.05)	0.233 (11.27)	-0.146 (-7.66)	0.287 (7.31)	-0.119 (-3.29)	0.107 (5.38)	--
<u>Type of lead vehicle (Other than motorcycle vehicle type is base)</u>								
Motorcycle	--	--	--	--	--	--	--	--
<u>Interaction between the motorcycle vehicle type and explanatory variable</u>								
Space gap in longitudinal direction (m)	-0.016 (-6.32)	--	-0.012 (-3.82)	--	-0.015 (-6.22)	--	-0.006 (-2.47)	--
Relative speed in longitudinal direction (m/s)	-0.065 (-2.96)	0.061 (2.75)	-0.053 (-4.34)	0.050 (5.25)	-0.054 (-2.49)	0.051 (2.34)	-0.031 (-2.66)	0.017 (1.48)
Stimuli from MF2 (second vehicle in MF)								
Subject vehicle has 2 or more lead vehicles (DM2) (One lead vehicle is base)	-0.469 (-4.05)	--	-0.261 (-2.02)	0.108 (2.67)	-0.446 (-4.22)	--	-0.045 (-1.09)	--
DM2 × Space gap in longitudinal direction (m)	0.017 (3.40)	--	0.006 (1.18)	--	0.015 (3.29)	--	--	--
DM2 × Relative speed in longitudinal direction (m/s)	0.108 (6.00)	-0.025 (-1.45)	0.086 (8.13)	--	0.096 (5.40)	-0.020 (-1.22)	0.048 (5.13)	--
Stimuli from LF1 (first vehicle in LF)								
Subject vehicle has 1 or more lead vehicles (DL1) (No lead vehicle is base)	--	0.266 (3.75)	-0.204 (-2.63)	--	--	0.260 (4.01)	-0.100 (-1.53)	--
DL1 × Space gap in longitudinal direction (m)	--	-0.009 (-3.44)	0.009 (3.01)	-0.004 (-1.37)	--	-0.010 (-4.27)	0.006 (2.47)	--
DL1 × Lateral gap between MF1 and LF1 (m)	--	-0.062 (-4.44)	0.060 (3.83)	-0.022 (-1.63)	--	-0.060 (-4.71)	0.035 (2.66)	--
DL1 × Relative speed in longitudinal direction (m/s)	0.063 (8.51)	--	0.034 (4.89)	-0.010 (-1.71)	0.058 (8.38)	--	0.011 (1.67)	--
Stimuli from RF1 (first vehicle in RF)								
Subject vehicle has 1 or more lead vehicles (DR1) (No lead vehicle is base)	--	0.148 (2.39)	--	0.155 (2.91)	--	0.121 (2.05)	-0.101 (-1.93)	0.127 (2.74)
DR1 × Space gap in longitudinal direction (m)	--	-0.007 (-2.24)	--	--	--	-0.006 (-2.27)	--	--
DR1 × Lateral gap between MF1 and RF1 (m)	--	--	0.022 (1.03)	--	--	--	0.036 (1.75)	--
DR1 × Relative speed in longitudinal direction (m/s)	0.037 (4.26)	--	0.030 (4.13)	-0.030 (-4.10)	0.031 (3.65)	--	0.020 (2.83)	-0.015 (-2.18)
Stimuli from LS1 (first vehicle in LS)								
Subject vehicle has 1 or more side vehicle (DLS1) (No side vehicle is base)	-0.092 (-1.60)	--	-0.164 (-2.51)	--	-0.078 (-1.42)	--	-0.132 (-2.39)	--
DLS1 × Space gap in lateral direction (m)	0.046 (2.74)	--	0.045 (2.46)	--	0.037 (2.29)	--	0.030 (1.91)	--
DLS1 × Relative speed in longitudinal direction (m/s)	--	-0.012 (-1.41)	0.033 (3.99)	--	--	-0.012 (-1.58)	0.023 (3.11)	--
Stimuli from RS1 (first vehicle in RS)								
Subject vehicle has 1 or more side vehicle (DRS1) (No side vehicle is base)	-0.064 (-1.44)	--	--	--	-0.063 (-1.54)	--	--	--
DRS1 × Space gap in lateral direction (m)	--	--	--	--	--	--	--	--
DRS1 × Relative speed in longitudinal direction (m/s)	--	--	--	-0.032 (-3.30)	--	--	--	-0.026 (-2.86)
Road geometry characteristics								
Space gap between left edge of the SV and left edge of the road (m)	--	-0.046 (-3.14)	0.062 (4.52)	-0.035 (-2.57)	--	-0.047 (-3.30)	0.021 (1.63)	-0.021 (-1.73)
Copula dependency parameter (θ)	NA	NA	NA	NA	NA	NA	-4.232 (-12.68)	-4.991 (-17.73)
Scale parameter	NA	NA	0.927 (48.74)	0.921 (52.59)	NA	NA	0.927 (48.74)	0.921 (52.59)
Goodness of fit measures								
Number of parameters	65				62			
Log likelihood	-24131.08				-24033.43			
BIC value	48898.51				48673.84			
Adjusted rho-square	0.13				0.14			
Number of cases	17852				17852			

Maintain same speed is base, *Dependent variable = absolute value of acceleration/deceleration at t s (m/s^2), -- the corresponding parameter was dropped from the specification as it was found to be statistically insignificant, NA- Not applicable.

Table 3.5 Estimation results for different number of lead vehicles in the model.

Goodness of fit measures of models considering influence of …	Influence of 1 lead vehicle in middle front compartment	Influence of 1 lead vehicle in middle front compartment by considering type of lead vehicle and interaction with other stimuli from same vehicle	Influence of 2 vehicles in middle front compartment*	Influence of 2 vehicles in middle front and 1 in left front compartment*	Influence of 2 vehicles in middle front, 1 in left front, 1 in right front compartment*	Influence of 2 vehicles in middle front, 1 in left front, 1 in right front, 1 on left side compartment*	Influence of 2 vehicles in middle front, 1 in left front, 1 in right front, 1 on left side, 1 on right side compartment*	Influence of 2 vehicles in middle front, 1 in left front, 1 in right front, 1 on left side, 1 on right side compartment and road geometry characteristics*
Model number	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Log likelihood	-24726.27	-24354.95	-24238.55	-24150.36	-24068.42	-24054.68	-24046.29	-24033.43
Number of significant parameters	20	29	35	43	51	57	59	62
Number of restrictions	--	9	6	8	8	6	2	3
LLR	--	742.64	232.79	176.38	163.88	27.49	16.78	25.71
χ^2 at 95 % confidence level	--	16.92	12.59	15.51	15.51	12.59	5.99	7.81
Remark	--	Model 1 is rejected compared to model 2	Model 2 is rejected compared to model 3	Model 3 is rejected compared to model 4	Model 4 is rejected compared to model 5	Model 5 is rejected compared to model 6	Model 6 is rejected compared to model 7	Model 7 is rejected compared to model 8
Number of cases	17852							

*Type of lead vehicle and interaction with other stimuli from same vehicle is also considered in the model specification

3.6 VALIDATION

A validation exercise was performed on a hold-out dataset of 5,000 vehicle manoeuvres to assess the efficacy of the proposed model in predicting driver's discrete and continuous decisions at an aggregate level. To do so, the following procedure was followed for simulating the driver's discrete and continuous decisions for each of the 5,000 data points $(q = 1, 2, ..., 5,000)$.

1. Simulate the discrete decisions: To simulate discrete decisions of acceleration, deceleration, or maintain same speed, we used the following two-step approach:

   (1a). Calculate the probability of each possible discrete choice decision ($i = a, d, s$) using the following multinomial logit expression.

$$P_{qi} = \frac{\exp\left(\beta_i^T x_{qi}\right)}{\sum_j \exp\left(\beta_j^T x_{qj}\right)};\ i = a, d, s \tag{3.22}$$

   (1b). Use the above calculated discrete choice probabilities to simulate the discrete decision of acceleration ($i = a$), deceleration ($i = d$), or maintain constant speed $(i = s)$. This can be done by drawing a pseudo random number from a uniform [0,1] distribution and seeing where the draw falls in the cumulative probability space of the discrete choices $(i = a, d, s)$.

2. For a given simulated discrete decision of acceleration or deceleration ($i = a, d$), simulate the corresponding continuous decision (acceleration and deceleration) values. To do so, compute the conditional expected value of the corresponding continuous decision (i.e., extent of acceleration or extent of deceleration) using the following expression:

$$E\left[f\left(m_{qi}|i\right)\right] = \int_{L_i}^{U_i} \left(f\left(m_{qi}|i\right) m_{qi} dm_{qi}\right);\ i = a, d \tag{3.23}$$

   where, $f\left(m_{qi}|i\right)$ is the conditional density of the corresponding continuous decision (acceleration and deceleration) and takes the following expression from Eq. (3.15), and all other terms are defined in Section 3.3:

$$f\left(m_{qi}|i\right)=\left(P_{qi}\right)^{-1}\left\{\frac{1}{\sigma_{\eta i}}f_{\eta i}\left(\frac{m_{qi}-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)-\frac{1}{\sigma_{\eta i}}f_{\eta i}\left(\frac{m_{qi}-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)\frac{\partial C_\theta\left(u_{q1}^i,u_{q2}^i\right)}{\partial u_2^i}\right\};\ i=a,d \quad (3.24)$$

Since the predicted (i.e., simulated) discrete decisions depend on the random draw used for simulation of the probabilities in Eq. (3.22), the above procedure was repeated for 100 simulated pseudo random draws for each of the 5,000 data points (i.e., a total of 500,000 simulations). Next, the percentage of the times each of the discrete decisions ($i=a,d,s$) was predicted from the above 500,000 simulations was computed. Subsequently, across all the simulations when acceleration was predicted, the average of the conditional expected value of acceleration was computed (i.e., average of the values computed using Eq. (3.23)). Similarly, across all the simulations when deceleration was predicted, average of the conditional expected value of deceleration was computed. These aggregated predictions are shown in Table 3.6, along with the corresponding aggregated observed values from the data. As can be seen from the table, the percentage of the times vehicles were observed to have accelerated, decelerated, or remained in constant speed is close to the percentage of the times vehicles were simulated to have taken those discrete choice decisions. Also, the aggregate predictions of the extents of acceleration and deceleration values are close to the corresponding observed values. Specifically, the observed average predicted acceleration (deceleration) value for the vehicles predicted to have accelerated (decelerated) is very close to that of vehicles that were observed to have accelerated (decelerated). These results suggest that the proposed model can predict aggregate patterns of the acceleration and deceleration decisions and the corresponding extents very well.

Table 3.6 Observed and predicted aggregate acceleration and deceleration decisions in a hold-out sample (N = 5,000 data records)

Discrete and continuous decisions	**Observed**	**Predicted**
Percentage of instances vehicles accelerated (discrete choice)	41.9%	41.3%
Average acceleration (average of conditional expected acceleration calculated using Eq. (3.23)) if the discrete decision is acceleration	0.75 m/s^2	0.74 m/s^2
Percentage of instances vehicles decelerated (discrete choice)	44.9%	46.5%
Average deceleration (average of conditional expected deceleration calculated using Eq. (3.23)) if the discrete decision is deceleration	0.71 m/s^2	0.72 m/s^2
Percentage of instances vehicles maintained same speed	13.2%	12.3%

3.7 CONCLUSIONS AND FUTURE RECOMMENDATIONS

This chapter proposes a multi-vehicle anticipation and copula-based discrete-continuous choice modelling framework for describing driver behaviour in HD traffic conditions. To incorporate the impact of multi-vehicle anticipation, we introduce the concept of influence zone around a subject vehicle and consider stimuli from all vehicles and other objects within the influence zone. Further, driving decisions are characterized as combination of discrete and continuous components. The discrete component involves the decision to accelerate, decelerate, or maintain a constant speed and the continuous component involves the decision of how much to accelerate or decelerate. The proposed copula-based joint modelling framework allows a flexible error dependency structure between driver's discrete and continuous decisions. Additionally, truncated distributions are employed for the continuous model components to avoid the prediction of unrealistically high acceleration or deceleration values.

The model estimation has been carried out using a vehicle trajectory dataset from Chennai, India. To capture the dependency structure between the driver's manoeuvring decision and its extent, various copula functions including Gaussian, FGM, Frank, Gumbel, Clayton, Joe, and AMH have been employed in this chapter. Overall, the Frank copula model with truncated distributions that consider bounds for acceleration and deceleration values performed better than all other models estimated in this chapter.

The empirical results of this chapter enhance the current understanding of driving behaviours in HD traffic conditions in several ways. First, the results demonstrate the importance of considering multi-vehicle anticipation for describing driving behaviour in HD traffic conditions. Specifically, the results lend credence to the idea of an influence zone around the vehicle and corroborate our hypothesis that driving behaviour is influenced by multiple vehicles within the influence zone of the subject vehicle. Further, drivers in HD traffic conditions not only consider vehicles that are ahead of their vehicle but also consider those vehicles that are on either side. At the same time, while the drivers' decision to accelerate and the extent of acceleration is governed by multiple vehicles ahead, their decision on the extent of deceleration is likely to be affected more by the immediate lead vehicle than other vehicles in the influence zone. Such nuanced findings have not been reported by other studies that considered multi-vehicle anticipation in driving behaviour.

Second, most of the earlier studies in the literature model driver behaviour as a single continuum, where the discrete decisions (of whether to accelerate, decelerate, or remain in constant speed) and continuous decisions (how much to accelerate or decelerate) are not treated separately. However, this chapter demonstrates that driving behaviour is more complex. The results support the notion that discrete and continuous decisions are separate and the factors influencing the discrete decisions might be different or may influence the continuous decisions differently. Specifically, not all traffic environment variables found to influence the discrete decisions were found influential on continuous decisions (or how much to accelerate or decelerate) and vice versa. Moreover, the influence of several variables was found to be stronger on the decision to accelerate or decelerate than on the decision of how much to accelerate or decelerate. This suggests that the discrete and continuous decisions likely require different cognitive efforts by drivers and are influenced in different ways by the traffic environment variables. Such findings add to our understanding of driver decision making process and driving task performance.

Third, the results indicate the presence of significant unobserved factors contributing to the negative dependency between the driver's discrete and continuous decisions. This suggests that the unobserved factors that increase (decrease) the propensity to accelerate/decelerate are more likely to decrease (increase) the extent of acceleration/deceleration. When compared to the estimation results of an independent model that ignores such dependencies, the copula-based joint models offer a statistically superior goodness-of-fit and more plausible behavioural interpretations. This highlights the importance of jointly modelling the discrete and continuous decisions. In addition, employing truncated normal distributions to model acceleration and deceleration values not only led to a significant improvement in the model fit but also prevents the possibility of unrealistically high acceleration or deceleration values.

CHAPTER 4 A PANEL DATA-BASED DISCRETE-CONTINUOUS MODELLING FRAMEWORK TO ANALYSE LONGITUDINAL DRIVER BEHAVIOUR IN HOMOGENEOUS AND HETEROGENEOUS DISORDERLY TRAFFIC CONDITIONS

Abstract

In this chapter, we propose a panel data-based discrete-continuous modelling framework to analyse driver behaviour in two disparate trajectory datasets – one from a heterogeneous, disorderly (HD) traffic stream in India and another from a homogeneous traffic stream in the United States. The panel data-based framework allows the analyst to isolate the subject vehicle- and driver-specific unobserved factors that influence driver behaviour. Doing so helps reduce the confounding effects of such unobserved factors on analysing the influence of observed factors, such as relative speeds and spacing between the subject vehicle and other vehicles, on driver behaviour. This also helps reduce the confounding effects of unobserved factors on analysing the differences in driving behaviour between HD and homogeneous traffic streams. The empirical results reveal both similarities and differences in driver behaviour between the two trajectory datasets examined in this chapter. Specifically, in both datasets, in addition to vehicles ahead of the subject vehicle in its lane, vehicles ahead in the adjacent lanes were found to influence driver behaviour. However, side vehicles were found to influence drivers' decision-making only in the HD traffic dataset. We also examined the suitability of different lengths of influence zones on drivers' longitudinal movement behaviour in both traffic datasets. In this context, a 60 m length influence zone was found more suitable than shorter length zones to model driver behaviour in the HD traffic trajectory dataset from India. In contrast, a 30 m length influence zone was found more suitable for the homogeneous traffic stream trajectory dataset in the United States. Such insights can help improve driver behaviour models and traffic simulation frameworks for both traffic conditions.

Note: The material in this chapter is drawn from the following paper:

Nirmale, S. K., Pinjari, A. R., and Sharma, A. (2022). A panel data-based discrete-continuous modelling framework to analyse longitudinal driver behaviour in homogeneous and heterogeneous disordered traffic conditions. *Transportation Letters*. https://doi.org/10.1080/19427867.2022.2132058

4.1 INTRODUCTION

In general, traffic flow conditions on uninterrupted traffic facilities across the world can be divided into two broad categories, namely, *homogeneous traffic* flow conditions and *heterogeneous, disorderly* (HD) traffic flow conditions. *Homogeneous traffic* flow is typically observed in countries such as Australia, the USA, Germany, and the Netherlands, whereas HD traffic flow is typically observed in Asian countries such as India, Bangladesh, and Indonesia. The characteristics of HD traffic flow tend to be significantly different from those of *homogeneous traffic* flow. Specifically, HD traffic streams comprise a wide variety of vehicle classes (such as passenger cars, motorcycles, buses, trucks, three-wheeled auto-rickshaws, and non-motorised vehicles) with considerably different physical and operational characteristics. Most of these classes have substantial representation in the traffic streams. In contrast, *homogeneous traffic* streams are dominated mostly by passenger cars having similar physical and operational characteristics. Moreover, contrary to *homogeneous traffic*, vehicles in HD traffic exhibit weak to no lane discipline, a greater extent of lateral movements, staggered following, and squeezing in gaps between vehicles (Asaithambi *et al.*, 2016). Despite the substantial differences in traffic conditions, most of the previous studies have investigated driver behaviour either only in HD traffic conditions (Chakroborty et al., 2004; Dey et al., 2008; Venkatachalam and Gnanavelu, 2009; Mallikarjuna and Rao, 2011; Mathew et al., 2013; Choudhury and Islam, 2016; Das and Maurya, 2018a; Kanagaraj and Treiber, 2018; Raju et al., 2018; Das et al., 2019; Sarkar et al., 2020; Amrutsamanvar et al., 2021; Raju et al., 2021) or in *homogeneous traffic* conditions (Gazis et al., 1961; Gipps, 1981; Bando et al., 1995; Ahmed, 1999; Treiber et al., 2000; Koutsopoulos and Farah, 2012). Limited efforts have been made to model driver behavior in both homogeneous and HD traffic streams using a same modelling framework.

To model the longitudinal movements of a subject vehicle, most available models consider stimulus from either a single lead vehicle only (Herman et al., 1959; Gazis et al., 1961; Bando et al., 1995; Treiber et al., 2000) or multiple vehicles within the same lane (Bexelius, 1968; Lenz et al., 1999; Hoogendoorn et al., 2006; Zhang, 2014). This is because such models are developed for *homogeneous traffic* conditions, where the driver's focus is typically on the vehicles ahead in the same lane unless the driver intends to change lanes. In HD traffic, however, due to a lack of lane discipline and a wide variety of vehicle classes, the subject vehicle might be behind more than one vehicle or in between multiple leaders. Therefore, as discussed in previous studies (Jin et al., 2010; Li et al., 2015; Choudhury and Islam, 2016; Li et al., 2016; Budhkar and Maurya, 2017), models that consider a single lead vehicle cannot truly represent the driver behaviour in HD traffic conditions. Further, models that consider stimuli from multiple vehicles in the same lane of the subject vehicle are also not suitable since Chapter 3 has demonstrated that side vehicles and vehicles present in the left front or right front locations with respect to a subject vehicle influence its driver's longitudinal movements in HD traffic conditions. Therefore, a model that considers the influence of vehicles that are not only straight ahead but also on the sides of the subject vehicle will be suitable to mimic longitudinal movements in HD traffic conditions. Intuitively, such a model can potentially be applied for modelling longitudinal movements in *homogeneous traffic* conditions as well since these conditions represent a special case of HD traffic conditions.

Another important point to note is that, in both traffic conditions, most existing models do not consider the influence of driver-specific unobserved factors such as age, gender, and aggressiveness on driver behaviour. Similarly, vehicle-specific factors such as the engine's kinematic capabilities might cause variations in acceleration and deceleration behaviours across different vehicles. These factors remain mostly unobserved because the vehicle trajectory data used for building, calibrating, and validating driver behaviour models typically come from videos or GPS devices that do not measure such information. However, it is evident from previous studies that driver-specific human factors affect driver behaviour. For example, increased driver response time was observed for older drivers (Edwards et al., 2003), and aggressive drivers were observed to opt for shorter time gaps (Sharma et al., 2020). Hence, accounting for the impact of unobserved, driver-specific factors can increase the realism of driver behaviour models. Besides, such unobserved heterogeneity, when present but ignored, can potentially confound the other parameter estimates of driver behaviour models. This can potentially lead to biased estimation and distorted inferences of the influence of typically examined observed factors such as relative speeds and space gaps on driver behaviour.

In this chapter, we propose a panel data-based discrete-continuous modelling framework to analyse driver behaviour in two disparate trajectory datasets – one from a *heterogeneous, disorderly* (HD) traffic stream in India and another from a *homogeneous traffic* stream in the United States. The panel data-based model in this chapter builds on a multi-vehicle anticipation (MVA) based discrete-continuous choice model proposed in Chapter 3 that characterizes longitudinal movements of the subject vehicle as a combination of discrete choice (whether to accelerate, decelerate, or maintain constant speed) and continuous choice (how much to accelerate or decelerate). The panel data-based framework allows the analyst to isolate the subject vehicle- and driver-specific unobserved factors (such as age and aggressiveness) that do not vary across different observations of a same vehicle but have an influence on the vehicle's driver behaviour. Doing so helps reduce the confounding effects of such unobserved factors on analysing the influence of observed factors (such as relative speeds and spacing between the subject vehicle and other vehicles) on driver behaviour. This also helps in reducing the confounding effects of unobserved factors on analysing the differences in driving behaviour between HD and *homogeneous traffic* streams. Furthermore, in Chapter 3, we did not consider the issue of serial correlations that may arise due to correlation across successive observations from the trajectory of a vehicle. Such serial correlations, if present (but are ignored by the analyst) can also confound the parameter estimates and inferences made from the empirical models. This chapter employs a simple empirical strategy to remove serial correlation (more on this in Section 4.2.1.2).

Given that *homogeneous traffic* streams observed in the US and HD traffic streams typically observed in India are distinctively different, it would be interesting and insightful to compare and contrast the driver behaviour in trajectory datasets from these two types of traffic streams. Literature is devoid of studies investigating and comparing driver behaviour in HD traffic and *homogeneous traffic* streams. To the best of the authors' knowledge, only one study compared driver behaviour in homogeneous and HD traffic conditions (Ravishankar and Mathew, 2011). However, the modified Gipps's model employed in this study lacked the aforementioned features specific to HD traffic conditions, such as MVA, and the consideration of vehicle- and driver-specific unobserved factors. Hence, we compare the influence of various factors on car driver behaviour in two trajectory datasets – one from India and one from the US. Knowledge of similarities and differences in driver between the two datasets will be useful in building behaviourally realistic and robust driver behaviour models. In addition, understanding the differences in driver behaviour between the two types of traffic streams might be useful in avoiding any pitfalls in understanding and simulating HD traffic flow using models developed for *homogeneous traffic* conditions.

It is worth noting here that a comparative analysis of driving behaviour between the two different types of traffic streams involves trajectory data from disparate locations. When driving behaviour patterns are compared between such disparate datasets, the differences in behaviour may be attributed to not only the differences between HD and *homogeneous traffic* streams but also several other possible reasons. The other reasons include, for example, differences in the driver population (demographics and driving culture), vehicle characteristics, and other factors such as left-side vs. right-side driving, geometric features of roadways, congestion levels, season and weather conditions, and time-of-day of data collection. Since many of these factors, such as driver population characteristics, are typically unobserved in trajectory datasets, it is useful to control for such unobserved effects in analyses involving a comparison of driving behaviour between datasets from disparate locations. The panel data models proposed in the chapter help control for the confounding effects of unobserved factors (e.g., subject vehicle- and driver-specific factors) that vary across different vehicles in each of the two datasets. However, the models do not help control for factors that are different between the two datasets but are not different across the vehicles within each dataset. These include, for example, congestion levels and traffic composition in the two traffic streams, roadway geometry, and type of facilities. To control for the effects of such factors when comparing driving behaviour between HD and *homogeneous traffic* conditions, it is important to analyse a greater variety of trajectory datasets from a larger number of locations representing variation in such factors in both HD traffic and *homogeneous traffic* conditions. Despite this limitation, we believe it is helpful to start documenting the differences observed in trajectory datasets from different geographies. More such studies in the future with additional datasets can help build a repository of findings that might lend themselves to a larger meta-study.

In the rest of this chapter, Section 4.2 discusses the modelling framework and elucidates how it is different from that in Chapter 3. Section 4.3 provides an overview of the trajectory datasets and model variables. Section 4.4 presents and discusses the model estimation results. Section 4.5 summarises the main findings of this chpater.

4.2 METHODOLOGY

4.2.1 Methodological Issues Considered

4.2.1.1 Influence zone and MVA

Following Chapter 3, we considered a rectangular shaped *'influence zone'* around each subject vehicle (SV), as shown in Figure 4.1[4]. The vehicles within this zone are considered to influence the SV's driver behaviour, as opposed to a single lead vehicle. For HD traffic conditions, since data come from a three-lane arterial, road boundaries on either side are approximated as the edges of a rectangular-shaped influence zone. For *homogeneous traffic* conditions, since data come from a six-lane highway, the sum of the width of immediate lanes on the left and right sides of the subject vehicle and the width of the current lane of the vehicle (i.e., a width of three

[4] Note that this figure from Chapter 3 is repeated here for the reader's convenience.

lanes) is considered as the width of the influence zone. For both traffic conditions, the rectangular influence zone is divided into the following five compartments: middle front (MF), left front (LF), right front (RF), right side (RS), and left side (LS) (refer to Figure 4.1). The labelling of surrounding vehicles in each of these compartments – MF1, MF2, LF1, RF1, LS1, and RS1 – is done as depicted in the legend of Figure 4.1.

We have also considered the impact of the third vehicle in the middle front compartment (MF3), the second vehicle in the left front compartment (LF2), the second vehicle in the left side compartment (LS2), and the second vehicle in right side compartment (RS2). Our initial estimation results indicated that these vehicles did not influence the driver's microscopic behaviour, and hence, these vehicles were not considered in subsequent analysis.

Next, we need to determine a suitable length of the influence zone for each of the two traffic streams. Literature suggests that vehicles in 30 m to 60 m range of an SV are likely to influence its driver behaviour (Herman and Potts, 1959; Subramanian, 1996; Panwai and Dia, 2007; Punzo et al., 2011; Kanagaraj et al., 2015; Sharma, Zheng, et al., 2018; Sarkar et al., 2020). In Chapter 3, we considered an influence zone length of 30 m in their empirical analysis of driver behaviour in HD traffic streams. However, not much guidance exists on what length of the influence zone is suitable for each of HD and *homogeneous traffic* conditions. In this context, the differences in the characteristics of the two traffic streams may lead to a difference in the lengths of influence zones considered by their respective drivers. Therefore, to better understand the differences in driver behaviour, it is worth investigating the appropriate size of the influence zones in the two traffic streams. To this end, in this chapter, we relax the assumption of a fixed-length influence zone and investigate three different lengths of influence zones – 30 m, 45 m, and 60 m – for both traffic streams.

Note here that, regardless of the length, the influence zone gets trimmed accordingly when the subject vehicle is at the extreme left or right edge of the road. For instance, the influence zone has only three compartments – MF, RF, and RS – when the SV is at the extreme left edge of the road. Further, the back edge of the influence zone follows the SV's back bumper (see Figure 4.1).

In this chapter, an SV's driver behaviour at any time instance is characterised as a combination of discrete and continuous choices. The discrete choice comprises the manoeuvring decision of whether to accelerate, decelerate, or maintain a constant speed. If the driver decides to accelerate (decelerate), the continuous choice involves the extent of acceleration (deceleration). This is because not all the factors influencing the discrete choices might influence the continuous choices or vice versa, and the cognitive efforts needed to make the decision of whether to accelerate or decelerate might be different from that of how much accelerate or decelerate. At the same time, drivers make both the discrete and continuous choices jointly, and several common unobserved factors specific to drivers, vehicles, and the traffic environment might influence both discrete and continuous decisions. Therefore, it is useful to model these decisions jointly while recognising the correlations (or dependencies) between the discrete and continuous choice modelling components. While these conjectures have been corroborated in the context of HD traffic conditions in Chapter 3, the current chapter examines the conjectures for *homogeneous traffic* conditions also.

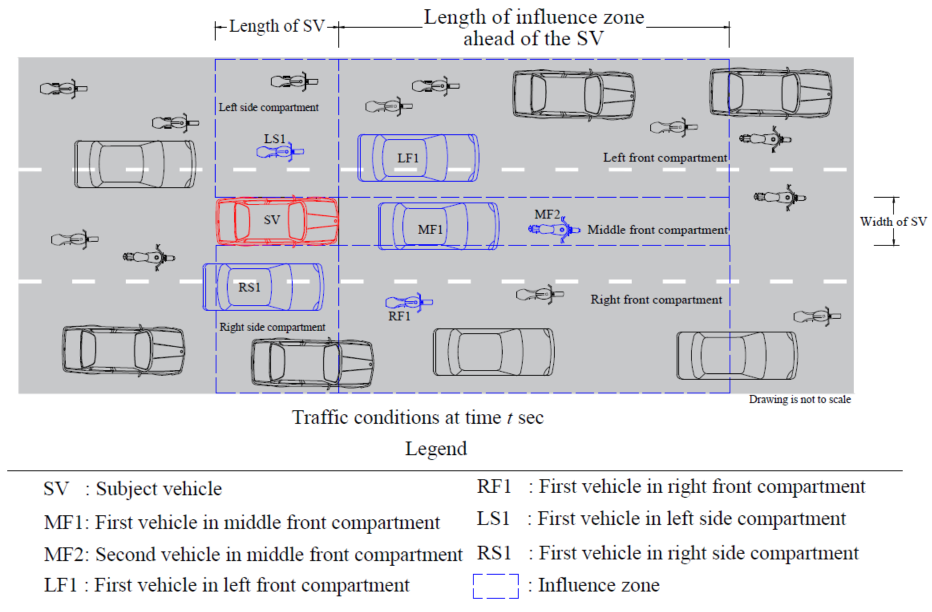


Figure 4.1 Structure of a rectangular influence zone around a subject vehicle

4.2.1.2 Serial correlation

Methodological considerations new in this chapter (not considered in Chapter 3) are two-fold: (a) serial correlation among successive observations (time series) of the same SV, and (b) SV- and driver-specific unobserved factors that influence driver behaviour and persist across all observations of a given SV (thereby necessitating the use of panel data models).

Serial correlation due to the similarity of the traffic environment between temporally proximate observations is unavoidable in time series data such as vehicle trajectory data. If such serial correlation exists and is ignored, the standard errors of the parameter estimates tend to be underestimated. As a result, the tendency to reject the null hypothesis for estimated parameters – when it should not be rejected – increases. To address the serial correlation issue in each of the HD and *homogeneous traffic* conditions datasets, we estimated linear regression models on the observed accelerations in the trajectory data (treating the values of acceleration, deceleration, or near-zero accelerations for vehicles in constant speed as arising from a single normal distribution) and calculated the residuals using these regression models. We used these residuals to conduct the widely used Durbin-Watson test for serial correlation. The Durbin-Watson test statistic (DW) for a given trajectory dataset is calculated using the following expression (Durbin and Watson, 1950):

$$DW = \frac{\sum_{q=1}^{Q} \sum_{qn=2}^{N_q} \left(e_{qn} - e_{qn-1}\right)^2}{\sum_{q=1}^{Q} \sum_{qn=2}^{N_q} \left(e_{qn}\right)^2} \tag{4.1}$$

where, e_{qn} is the residual from the regression equation for the qn^{th} data point in a time series of trajectory data for a subject vehicle q; N_q is the number of data points in the time series for the subject vehicle q; and Q is the number of vehicles. The hypotheses generally considered in the Durbin-Watson test are as follows – null hypothesis (H_0): zero serial correlation in the residuals, alternative hypothesis (H_A): residuals have a serial correlation. The upper critical values (dw_u) and lower critical values (dw_l) used to conduct the test are available from standard statistics books (Gujarati et al., 2012; Wooldridge, 2012) for different values of significance level and the number of explanatory variables (degrees of freedom) in the regression model. The possible outcomes from the Durbin-Watson test are the following: do not reject H_0 if $DW > dw_u$ (i.e., serial correlation is insignificant); reject H_0 if $DW < dw_l$ (i.e., serial correlation is significant); and inconclusive test if $dw_l \leq DW \leq dw_u$.

When we applied the Durbin-Watson test for the trajectory data that retained observations of all time steps of each subject vehicle, we found that $DW < dw_l$ at 95 % confidence level in both HD and homogeneous datasets, indicating that the residuals of data from one time step were correlated to those form the subsequent time steps (i.e., the observed acceleration data demonstrated serial correlation). Therefore, we selected the data points from the time series of each vehicle, removing the consecutive data points in the time series until the residuals were found to be not correlated according to the Durbin-Watson test. In both HD and homogeneous datasets, we found that serial correlation was not significant when we selected data points that were at least 2.5 s apart from each other for each subject vehicle.[5] Since the model parameters are estimated using individual data points in the trajectory (as opposed to complete trajectories), this simple approach obviates the need for sophisticated methods to deal with serial correlation.

4.2.1.3 Panel effects

Even after removing serial correlation, SV- and its driver-specific unobserved factors (such as vehicle engine performance and driver characteristics) that influence the SV's microscopic movements might persist over time and cause correlation across the different observations of an SV. Such unobserved factors necessitate the use of panel data modelling approaches, as opposed to the cross-sectional data models used in most other driver behaviour studies that employ statistical models on vehicle trajectory data. It is worth noting that such panel data effects due to SV- and driver-specific unobserved factors might persist even after accounting for dependencies between the discrete and continuous choice components due to common unobserved factors that vary across the different observations of an SV. The current chapter accommodates dependencies due to unobserved effects at both levels – (a) dependencies between discrete and continuous choice components at the level of each observation (at a given time instance) of an SV and (b) panel effects due to unobserved factors at the level of the SV. A copula-based method is used for the former, and random effects are specified for the latter, as discussed below.

4.2.2 Joint Discrete-Continuous Modelling Framework: Formulation and Estimation

Let q denote a subject vehicle (SV) and let i denote its driver's manoeuvring choice alternatives ($a=$ accelerate, $d=$decelerate, $s=$maintain constant speed) at a time instance t. Using these notational preliminaries, consider the following set of equations used to model the discrete and continuous choices of drivers:

$$u_{qit}^{*} = \beta_i^T x_{qit} + \omega_{qi} + \psi_{qi} + \varepsilon_{qit};\ i = a, d, s \tag{4.2}$$

$$.\, m_{qit} = \alpha_i^T z_{qit} \pm \omega_{qi} + \xi_{qi} + \eta_{qit};\ i = a, d\ . \tag{4.3}$$

[5] When all datapoints in the trajectories were considered, the DW statistic values for the HD and homogeneous datasets were 1.335 and 0.502, respectively, which were smaller than the corresponding lower critical value (1.957). When we selected data points that were at least 2.5 s apart from each other for each subject vehicle, the DW statistic values for the HD and homogeneous datasets were 2.0010 and 2.140, respectively, which were greater than the corresponding upper critical value (1.966).

In Eq. (4.2) above, u_{qit}^{*} is the utility that the driver of the q^{th} vehicle perceives from choosing a manoeuvring decision i at time t; x_{qit} is a column vector of observed traffic environment variables that influence the utility for discrete decision i and β_i is the corresponding coefficient vector. Eq. (4.3) represents regression equations for m_{qit}, the extent of acceleration or deceleration conditional on whether the driver decides to accelerate ($i = a$) or decelerate ($i = d$). m_{qit} is mapped to observed traffic environment variables (z_{qit}) influencing the extent of acceleration or deceleration through the coefficient vector α_i.

The remaining terms in Eq. (4.2) and Eq. (4.3) are random error components. Among these, ω_{qi}, ψ_{qi}, and ξ_{qi} are normally distributed error components with mean zero and variances $\sigma_{i\omega}^2$, $\sigma_{i\psi}^2$, and $\sigma_{i\xi}^2$, respectively. These three error components represent the SV- and driver-specific unobserved factors that do not vary across the different observations of an SV and influence the decisions of the driver. The ω_{qi} terms ($i = a,d$) capture such SV- and driver-specific factors that influence both the discrete and continuous decisions, the ψ_{qi} terms $(i = a,d,s)$ capture the factors that influence the discrete choice decisions only, and ξ_{qi} $(i = a,d)$ capture the factors that influence the continuous choices only. Note that the ω_{qi} terms which enter Eq. (4.2) also enter Eq. (4.3). The same ω_{qi} terms enter the two sets of equations because these capture the SV- and driver-specific unobserved factors that influence both the discrete and continuous decisions (i.e., these factors are common to both the discrete and continuous choice equations). For example, driver aggressiveness, age, gender, etc., may be such unobserved factors that influence both discrete and continuous choices and cause a correlation between the corresponding equations. These ω_{qi} terms enter with a '$\pm$' sign in Eq. (4.3) because the underlying correlation between the discrete and continuous components due to unobserved factors ω_{qi} may be +ve or –ve. The specific sign can be decided empirically. A '$+$' sign implies that the unobserved factors that increase (decrease) the propensity for taking a discrete decision i also increase (decrease) the extent of that decision. On the other hand, a '$-$' sign implies that the unobserved factors ω_{qi} that increase (decrease) the propensity for taking a discrete decision i also decrease (increase) the extent of that decision. Finally, ε_{qit} $(i = a,d,s)$ and η_{qit} $(i = a,d)$ represent unobserved factors assumed to be independent and identically distributed across the manoeuvring choice alternatives and across SVs. The ε_{qit} terms are assumed to be standard Gumbel distributed and the η_{qit} terms are assumed to be normal distributed with mean zero and variance $\sigma_{i\eta}^2$.

Recall that the ω_{qi} terms appear in both Equations (4.2) and (4.3) to generate dependencies between the discrete and continuous choice components of a driver, due to unobserved factors that do not vary across observations of an SV. In addition to these SV- or driver-specific effects, ε_{qit} and η_{qit} could be correlated because of correlations between time-varying factors influencing the discrete and continuous choices. To capture such dependencies between ε_{qit} and η_{qit}, we employed copulas as in Chapter 3. For more details on Copula functions, refer to Bhat and Eluru (2009).[6]

We assume that a driver of the subject vehicle q chooses a manoeuvring decision i that offers the maximum utility among all manoeuvring decisions ($j = a,d,s$), as expressed below:

$$\begin{aligned} u_{qit}^* &> \max_{j=a,d,s,j\neq i} u_{qjt}^* \\ \beta_i^T x_{qit} + \omega_{qi} + \psi_{qi} + \varepsilon_{qit} &> \max_{j=a,d,s,j\neq i} u_{qjt}^* \\ \varepsilon_{qit} - \left\{ \max_{j=a,d,s,j\neq i} u_{qjt}^* \right\} &> -\beta_i^T x_{qit} - \omega_{qi} - \psi_{qi} \\ v_{qit} &> -\beta_i^T x_{qit} - \omega_{qi} - \psi_{qi} \end{aligned} \tag{4.4}$$

$$\text{where, } v_{qit} = \varepsilon_{qit} - \left\{ \max_{j=a,d,s,j\neq i} u_{qjt}^* \right\}$$

Eq. (4.3) represents a multinomial discrete choice model. Conditional on the error components ω_{qi} and ψ_{qi}, since the ε_{qit} terms are assumed as IID Gumbel distributed, it results in a multinomial logit (MNL) conditional likelihood expression for the discrete choice component.

[6] As an alternative to couplas, we explored introducing time varying common error components in Eq. (4.2) and Eq. (4.3) to generate dependencies between discrete and continuous model components. Such an approach leads to likelihood functions that involve computationally intensive bi-level integrals (see Bhat and Castelar, 2002) due to the presence of error components at two different levels – (1) one set of error components at the observation level to represent unobserved factors that vary across time instances and (2) another set of error components at the SV level to represent unobserved factors that do not vary across time instances. However, we found that using the copula functions to capture correlations at the observation level and common random error coponents to capture SV-level effects resulted in a model that was much less computationally intensive (because one level of simulation-based integration is obviated by the use of copulas) and yielded better goodness of fit to data.

To estimate the model parameters, let c_i be the parameter vector that stacks the β_i and α_i vectors. The individual model components described earlier are brought together, and the joint likelihood expression of all observations of an SV q over different time instances $t=1,2,...,T$ conditional on ω_{qi}, ψ_{qi}, and ξ_{qi} is provided below:

$$L_q\left(c_i \mid \omega_{qi}, \psi_{qi}, \xi_{qi}\right) = \prod_{t=1}^{T}\left[\prod_{i=1}^{I}\begin{Bmatrix} P\left(v_{qit} > -\left(\beta_i^T x_{qit} + \omega_{qi} + \psi_{qi}\right)\right)\times \\ P\left(m_{qit} | v_{qit} > -\left(\beta_i^T x_{qit} + \omega_{qi} + \psi_{qi}\right)\right) \end{Bmatrix}^{\delta_{qit}}\right] \quad \forall i \neq j \tag{4.5}$$

where, $\delta_{qit} = 1$ if the driver of the q^{th} vehicle chooses an alternative i at time t, otherwise $\delta_{qit} = 0$. I represents the total number of discrete choice alternatives. Given the assumptions made earlier in the discussion, Eq. (4.5) can be expressed as below:

$$L_q\left(c_i \mid \omega_{qi}, \psi_{qi}, \xi_{qi}\right) = \prod_{t=1}^{T}\left[\prod_{i=1}^{I}\left\{\frac{1}{\sigma_{\eta i}}\phi_{\eta i}\left(\frac{m_{qit} - \left(\alpha_i^T z_{qit} \pm \omega_{qi} + \xi_{qi}\right)}{\sigma_{\eta i}}\right)\left(1 - \frac{\partial C_\theta\left(u_{qt1}^i, u_{qt2}^i\right)}{\partial u_{qt2}^i}\right)\right\}^{\delta_{qi}}\right] \tag{4.6}$$

In the above expression, C_θ is a copula function, $\phi_{\eta i}(.)$ is the standard normal probability density function. u_{qt1}^i is expressed as below (since the discrete choice component, conditional on ω_{qi} and ψ_{qi}, takes a multinomial logit form):

$$u_{qt1}^i = 1 - \frac{\exp\left(\beta_i^T x_{qit} + \omega_{qi} + \psi_{qi}\right)}{\exp\left(\beta_i^T x_{qit} + \omega_{qi} + \psi_{qi}\right) + \exp\left(\ln\sum_{j\neq i}\exp\left(\beta_j^T x_{qjt} + \omega_{qj} + \psi_{qj}\right)\right)}, \tag{4.7}$$

and u_{q2}^i is expressed as below:

$$u_{qt2}^i = \Phi\left(\frac{m_{qit} - \left(\alpha_i^T z_{qit} \pm \omega_{qi} + \xi_{qi}\right)}{\sigma_{\eta i}}\right), \tag{4.8}$$

where, $\Phi(.)$ is a standard normal cumulative density function. Integrating the conditional likelihood $L_q\left(c_i \mid \omega_{qi}, \psi_{qi}, \xi_{qi}\right)$ over the distributions of ω_{qi}, ψ_{qi} and ξ_{qi} results in the following unconditional likelihood for a driver of the vehicle q:

$$L_q\left(c_i,\sigma_{i\omega}^2,\sigma_{i\psi}^2,\sigma_{i\xi}^2\right)=\int_{\omega_{qi},\psi_{qi},\xi_{qi}} L_q\left(c_i \mid \omega_{qi},\psi_{qi},\xi_{qi}\right)\phi\left(\omega_i;\sigma_{i\omega}^2\right)\phi\left(\psi_{qi};\sigma_{i\psi}^2\right)\phi\left(\xi_{qi};\sigma_{i\xi}^2\right)d\omega_i d\psi_i d\xi_i \tag{4.9}$$

Note that the vector c_i , $\sigma_{i\omega}^2$, $\sigma_{i\psi}^2$, and $\sigma_{i\xi}^2$ are the population-level parameters to be estimated. The multivariate integral in the likelihood function of Eq. (4.9) may be simulated to result in the following simulated likelihood function as an estimator of $L_q\left(c_i,\sigma_{i\omega}^2,\sigma_{i\psi}^2,\sigma_{i\xi}^2\right)$:

$$SL_q\left(c_i,\omega_{qi},\psi_{qi},\xi_{qi}\right)=\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T}\left[\prod_{i=1}^{I}\left\{\frac{1}{\sigma_{\eta i}}\phi_{\eta i}\left(\frac{m_{qit}-\left(\alpha_i^T z_{qit}\pm\omega_{qi}^r+\xi_{qi}^r\right)}{\sigma_{\eta i}}\right)\left(1-\frac{\partial C_\theta\left(u_{qt1}^{ir},u_{qt2}^{ir}\right)}{\partial u_{qt2}^{ir}}\right)\right\}^{\delta_{qi}}\right]\right) \tag{4.10}$$

where, ω_{qi}^r, ψ_{qi}^r, and ξ_{qi}^r are the r^{th} draws from the distribution of ω_{qi}, ψ_{qi}, and ξ_{qi}, respectively, and R is the total number of draws covering the distributions. Note that ψ_{qi}^r is not visible in Eq. (4.10), but it enters through the u_{qt1}^i terms. The simulated log-likelihood expression for a sample of Q subject vehicles $(q=1,2,\ldots,Q)$ may be written as:

$$SLL\left(c_i,\sigma_{i\omega}^2,\sigma_{i\psi}^2,\sigma_{i\xi}^2\right)=\sum_{q=1}^{Q}\ln\left(SL_q\left(c_i,\omega_{qi},\psi_{qi},\xi_{qi}\right)\right) \tag{4.11}$$

The above simulated log-likelihood function is maximised to estimate the model parameters $\left(c_i,\sigma_{i\omega}^2,\sigma_{i\psi}^2,\sigma_{i\xi}^2\right)$. In this chapter, we used R = 500 Halton draws (Bhat, 2003) to simulate the distributions of $\{\omega_{qi},\psi_{qi},\xi_{qi}\}$ for computing $SLL\left(c_i,\sigma_{i\omega}^2,\sigma_{i\psi}^2,\sigma_{i\xi}^2\right)$.

4.3 DATASETS AND VARIABLES CONSIDERED IN THE MODEL

4.3.1 Empirical Datasets

The proposed discrete-continuous modelling framework was applied to two different datasets – one from a *homogeneous traffic* stream and another from an HD traffic stream. For *homogeneous traffic* trajectory data, we used the reconstructed NGSIM I80-1 dataset (Punzo et al., 2011; Montanino and Punzo, 2013; Montanino and Punzo, 2015). This NGSIM data were collected on eastbound I80 in the San Francisco Bay area in Emeryville, California. The study area was approximately 500 m in length and consisted of six freeway lanes. For HD traffic conditions, we used a vehicular trajectory dataset from an urban arterial stretch of 245 m in Chennai, India, originally processed by Kanagaraj et al. (2015). Both trajectory datasets include time series of vehicle positions (longitudinal and lateral), speeds, and acceleration/deceleration values. The time resolution in NGSIM and Chennai datasets is 0.1 s and 0.5 s, respectively.

To analyse a driver's decisions at a time instance t, we considered a variety of traffic environment variables at $t-\tau$ s as independent variables representing the potential stimuli where τ is the update time. To determine τ, exploratory linear regression models of the observed acceleration values as a function of traffic environment variables were estimated considering different update time values ranging from 0.1 s to 2.0 s. This chapter adopts the update time of 0.7 s and 0.5 s for *homogeneous traffic* conditions and HD traffic conditions, respectively, since these resulted in the best model fit and the most intuitive interpretation of the parameter estimates for the exploratory regression models.

4.3.2 Descriptive Statistics

For each of the two traffic conditions, we considered the same observations of the subject vehicles across different datasets created using 30 m, 45 m, and 60 m of influence zone lengths ahead of the subject vehicle. As a result, data on the dependent variables do not differ across these three datasets, but the explanatory variables do. This allows us to statistically compare the model estimated on these datasets. Table 4.1 provides key descriptive statistics of the processed trajectory datasets. The first set of descriptives corresponds to the percentages of instances vehicles have been observed to be in acceleration, deceleration, and constant speed states. Note that most trajectory datasets do not show a probability mass of exactly zero acceleration (to identify the percentage of constant speed instances). Therefore, following Chapter 3, we considered vehicles to be in a constant speed state if their acceleration value was in the range [-0.1 m/s^2, +0.1 m/s^2]. We also considered Ozaki's (1993) definition that a vehicle would be in a constant speed state when its acceleration is within [-0.05g, +0.05g], where g = 9.8 m/s^2. However, exploratory empirical models using both datasets offered much better intuitive interpretations with the former definition. Therefore, the former definition was retained for all subsequent analyses.

According to the above-discussed definition, the percentages of acceleration, deceleration, and constant speed state in the NGSIM dataset are 39.0 %, 37.6 %, and 23.4 %, respectively. The corresponding percentages in the Chennai dataset are 41.1 %, 46.3 %, and 12.6 %. Clearly, the percentage of the times vehicles were observed to be in a constant speed state is higher in *homogeneous traffic* conditions. This is perhaps because of orderly traffic due to lane discipline, lower rates of lateral movements, and greater space availability due to six freeway lanes in the *homogeneous traffic* situation (as opposed to relatively less orderly traffic on an urban arterial of only three lanes in the HD traffic situation). The average acceleration values (when vehicles accelerated) and deceleration values (when vehicles decelerated) are also higher in the *homogeneous traffic* dataset than those in the heterogeneous dataset. The standard deviations of observed acceleration and deceleration values, however, are smaller in the *homogeneous traffic* dataset. As can be observed from the table, in most instances, the subject vehicle in HD traffic conditions has at least one vehicle in the left front compartment, possibly because cars in India tend to travel on the right side of the road. Also, note that the average longitudinal speed of the subject vehicle, longitudinal space gap, and longitudinal speed difference with respect to MF1 vehicle is higher in the *homogeneous traffic* conditions than in HD traffic conditions. All these differences will likely have a bearing on the model estimation results.

Note that the mean space gaps with respect to LF1 are higher in HD traffic streams than those in *homogeneous traffic* streams, whereas mean space gaps with respect to RF1 are lower in HD traffic streams than in *homogeneous traffic* steams. This could be because of the differences in traffic rules followed in India and USA. In India, fast-moving vehicles travel on the right side of the road, whereas in the USA, they travel on the left side. As a result, traffic volumes in the right lanes are likely to be greater than traffic volumes in the left lanes in the Indian context (assuming that the speeds are not near the free-flow regime). Therefore, in the Indian context, the space gap of a subject vehicle with respect to a vehicle in the RF1 position is likely to be smaller than that in LF1.

Furthermore, generally, lateral gaps between vehicles are expected to be smaller for HD traffic streams than for homogenous traffic streams. This trend can be observed in the reported descriptive statistics. However, the mean lateral gap between MF1 and LF1 and that between MF1 and RF1 in HD traffic is only slightly smaller than the mean lateral gap in homogenous traffic stream (when a car is a subject vehicle). This may be due to the low traffic volume and low density observed in the Chennai trajectory dataset, i.e., only 6010 vehicles/hr and 370 vehicles/km for a three-lane urban arterial road. As the volume and density are low, higher lateral gaps among vehicles may be observed than what may be expected in more congested HD traffic streams.

4.4 MODEL ESTIMATION RESULTS AND DISCUSSION

The following statistical specifications were explored in this chapter on both traffic conditions: (a) independent discrete and continuous choice models without any dependencies, (b) joint discrete-continuous choice models with different types of copulas – Frank, Gaussian, FGM, Gumbel, Clayton, Joe, and AMH – to capture dependencies at the observation level, without considering panel effects or dependencies due to driver-level unobserved factors, (c) joint discrete-continuous choice models with copulas to capture observation-level dependencies and random error components to consider panel effects and dependencies due to driver-level unobserved factors. In addition, three different lengths of influence zones – 30 m, 45 m, and 60 m – were also considered for both datasets. The best fit models were decided by comparing goodness of fit measures such as AIC, BIC, and adjusted rho-squared values. In addition, the log-likelihood ratio test was employed whenever suitable.

Among all the models examined, the model with a Frank copula specification, driver-level unobserved effects, and 30 m length of the influence zone provided the best fit to the *homogeneous traffic* dataset. Whereas the model with a Frank copula specification, driver-level unobserved effects, and 60 m length of the influence zone provided the best fit to the HD traffic dataset. Note that while estimating the models, we considered explanatory variables (such as space gaps, relative speeds, and lateral gaps) corresponding to all vehicles present in a subject vehicle's influence zone. We found that models that considered explanatory variables with respect to two vehicles in the middle front compartment (MF1 and MF2), one in each of the left front and right front compartments (LF1 and RF1), and one on either side of the subject vehicle (LS1 and RS1) resulted in the most behaviourally consistent parameter estimates (refer to Figure 4.1 for the meanings of vehicle labels such as LF1 and RF1). Table 4.2 provides the parameter estimates of the final empirical specifications (from the standpoint of data fit and behavioural consistency) for both traffic conditions – considering a 30 m influence zone for the *homogeneous traffic* dataset and a 60 m influence zone for the HD traffic dataset. Table B.1 and Table B.2 in Appendix C provide parameter estimates corresponding to the other influence zones for both traffic streams. In addition, we compared the panel data-based discrete-continuous model developed in this chapter with the model developed in Chapter 3 using goodness of fit measures such as AIC, BIC, and adjusted rho-squared values. Table B.3 in Appendix B reports the goodness of fit measures for both the models. An examination of this table demonstrates that the panel data-based discrete-continuous model performs better than the model developed in Chapter 3.

Table 4.1 Descriptive statistics of explanatory variables corresponding to different lengths of influence zones for both homogeneous and HD traffic datasets*.

Variables	Homogeneous traffic conditions: Length of the influence zone ahead of the subject vehicle			Heterogeneous disorderly traffic conditions: Length of the influence zone ahead of the subject vehicle		
	30 m	**45 m**	**60 m**	**30 m**	**45 m**	**60 m**
<u>Discrete variables</u>	Percentage	Percentage	Percentage	Percentage	Percentage	Percentage
Dependent variable at *t* s						
Acceleration decision	39.0	39.0	39.0	41.1	41.1	41.1
Deceleration decision	37.6	37.6	37.6	46.3	46.3	46.3
Maintain same speed decision	23.4	23.4	23.4	12.6	12.6	12.6
Vehicles in different compartments of the influence zone						
Subject vehicle has one or more vehicles in the MF compartment	100	100	100	100	100	100
Subject vehicle has two or more vehicles in the MF compartment	40.0	75.1	86.8	38.0	65.8	77.9
Subject vehicle has one or more vehicles in the LF compartment	78.0	83.8	86.0	88.2	92.7	94.3
Subject vehicle has one or more vehicles in the RF compartment	73.7	79.2	81.0	49.0	54.7	57.2
Subject vehicle has one or more vehicles in the LS compartment	38.5	38.5	38.5	47.3	47.3	47.3
Subject vehicle has one or more vehicles in the RS compartment	38.0	38.0	38.0	22.2	22.2	22.2
<u>Continuous variables</u>	Mean (Std. dev.)	Mean (Std. dev.)	Mean (Std. dev.)	Mean (Std. dev.)	Mean (Std. dev.)	Mean (Std. dev.)
Dependent variable at *t* s						
Acceleration (m/s^2)	0.81 (0.63)	0.81 (0.63)	0.81 (0.63)	0.13 (0.99)	0.13 (0.99)	0.13 (0.99)
Deceleration (m/s^2)	-0.87 (0.74)	-0.87 (0.74)	-0.87 (0.74)	-0.17 (0.95)	-0.17 (0.95)	-0.17 (0.95)
Subject vehicle (SV) characteristics at $t-\tau$ s						
Speed in longitudinal direction (m/s)	7.82 (2.91)	7.82 (2.91)	7.82 (2.91)	6.07 (1.28)	6.07 (1.28)	6.07 (1.28)
Stimuli from MF1 (first vehicle in MF) at $t-\tau$ s						
Space gap in longitudinal direction (m)	13.63 (6.40)	13.63 (6.40)	13.63 (6.40)	13.87 (7.30)	13.87 (7.30)	13.87 (7.30)
Relative speed in longitudinal direction (m/s)	0.01 (1.38)	0.01 (1.38)	0.01 (1.38)	0.01 (1.42)	0.01 (1.42)	0.01 (1.42)
Stimuli from MF2 (second vehicle in MF) at $t-\tau$ s						
Space gap in longitudinal direction (m)	22.83 (4.81)	29.26 (8.21)	32.15 (10.70)	20.89 (6.03)	27.57 (9.47)	31.26 (12.35)
Relative speed in longitudinal direction (m/s)	0.09 (1.70)	0.09 (1.82)	0.10 (1.89)	0.07 (1.49)	0.02 (1.50)	0.01 (1.52)
Stimuli from LF1 (first vehicle in LF) at $t-\tau$ s						
Space gap in longitudinal direction (m)	10.68 (7.39)	12.46 (9.73)	13.46 (11.43)	8.61 (6.89)	9.92 (8.97)	10.64 (10.41)
Lateral gap between MF1 and LF1 (m)	2.03 (1.13)	2.02 (1.11)	2.02 (1.11)	2.21 (1.36)	2.20 (1.36)	2.19 (1.36)
Relative speed in longitudinal direction (m/s)	1.72 (4.50)	1.98 (4.72)	2.14 (4.86)	-0.47 (1.58)	-0.47 (1.59)	-0.47 (1.59)
Stimuli from RF1 (first vehicle in RF) at $t-\tau$ s						
Space gap in longitudinal direction (m)	10.39 (7.27)	12.19 (9.69)	13.06 (11.21)	10.89 (8.12)	13.52 (10.99)	15.23 (13.37)
Lateral gap between MF1 and RF1 (m)	1.50 (1.28)	1.63 (1.25)	1.68 (1.23)	1.67 (1.37)	1.80 (1.37)	1.87 (1.37)
Relative speed in longitudinal direction (m/s)	-0.13 (3.40)	-0.04 (3.41)	-0.01 (3.42)	0.30 (1.65)	0.28 (1.68)	0.25 (1.70)
Stimuli from LS1 (first vehicle in LS) at $t-\tau$ s						
Space gap in lateral direction (m)	2.02 (0.96)	2.02 (0.96)	2.02 (0.96)	2.17 (1.17)	2.17 (1.17)	2.17 (1.17)
Relative speed in longitudinal direction (m/s)	0.99 (3.96)	0.99 (3.96)	0.99 (3.96)	-0.48 (1.70)	-0.48 (1.71)	-0.48 (1.70)
Stimuli from RS1 (first vehicle in RS) at $t-\tau$ s						
Space gap in lateral direction (m)	2.00 (0.99)	2.00 (0.99)	2.00 (0.99)	1.64 (0.94)	1.64 (0.94)	1.64 (0.94)
Relative speed in longitudinal direction (m/s)	-0.33 (3.37)	-0.33 (3.37)	-0.33 (3.37)	0.24 (1.63)	0.24 (1.63)	0.24 (1.63)
Road geometry characteristics at $t-\tau$ s						
Space gap between left edge of SV and left edge of the road (m)	11.02 (6.06)	11.02 (6.06)	11.02 (6.06)	6.30 (1.89)	6.30 (1.89)	6.30 (1.89)
Number of vehicles		522			760	
Number of observations		8728			6914	

*Mean and standard deviation of variables with respect to surrounding vehicles are calculated when vehicles are present in the respective compartment.

4.4.1 Similarities in Statistical Specifications

For the dependencies between discrete and continuous choice decisions at the observation level, the copula parameter estimates had a negative sign in both datasets. This result suggests that the time-varying unobserved factors that increase (decrease) a driver's tendency to accelerate or decelerate also decrease (increase) driver's tendency to accelerate more or decelerate more. Although the same finding was reported and explained in Chapter 3 for HD traffic dataset, this chapter reveals negative dependencies in the *homogeneous traffic* dataset as well.

For the dependency due to unobserved factors at the driver level, we explored both positive and negative correlations and found that the model with positive correlations yielded a much better fit in both datasets. This implies that the driver-level unobserved factors that increase (decrease) a driver's propensity to accelerate also increase (decrease) the extent of acceleration by the driver. For instance, careful drivers are less likely to accelerate frequently, and they also accelerate less (i.e., do not accelerate more) when they accelerate.

The scale parameter estimates of the random error components ψ_{qi} that capture panel effects due to driver-specific unobserved factors on discrete decisions separately are statistically significant in both datasets. However, the scale parameter estimates of the ξ_{qi} terms that capture panel effects on continuous decisions (extent of acceleration or deceleration) separately are not significant in either dataset. This suggests that most of the driver-specific unobserved factors influencing the continuous decisions (extent of acceleration or deceleration) are already captured in the random error components (ω_{qi}) common to discrete and continuous decisions.

It is worth noting here that independent discrete and continuous choice models, which ignored the above-discussed dependencies and panel effects, resulted in inflated t-statistics and overestimated the influence of different traffic environment variables in both datasets. The consistency of these results between two different datasets from disparate traffic conditions suggests that statistical specification is an important issue in the context of parameter estimation (aka, calibration) of driver behaviour models. Since most vehicle trajectory datasets pose a panel data setting, the analyst cannot ignore the role of unobserved factors at either the observation level or the driver level.

4.4.2 Similarities in Driver Behaviour

The empirical models on both datasets support the stimulus-response theory used in most driver behaviour models in those stimuli such as relative speeds and spacing with respect to lead vehicles ahead (in the same lane) of the SV influence its driver's decisions. Specifically, the number of statistically significant parameters and the corresponding t-statistics indicate that for both models, the first lead vehicle (MF1) ahead of SV has the strongest influence on SV's driver behaviour, followed by the second lead vehicle (MF2) in the same lane except the acceleration decision case in HD traffic condition. Another interesting similarity is that vehicles in lanes adjacent to that of the SV – at least those in the LF and RF compartments – influence driver behaviour in both traffic conditions. This finding challenges a tacit assumption made by previous studies that SVs respond to stimuli from leaders in the same lane only. Further, in both datasets, not all factors that influence the discrete decisions influence the continuous decisions and vice versa. This finding supports the conjecture that the factors considered and cognitive efforts involved in deciding whether to accelerate, decelerate, or maintain constant speed are not necessarily the same as those in deciding how much to accelerate or decelerate.

4.4.3 Differences in Driver Behaviour

4.4.3.1 Differences in the size of the influence zone

As mentioned earlier, influence zones of lengths 30 m and 60 m provided the best fit for homogeneous and HD traffic datasets, respectively. This can be observed in all metrics used for the evaluation of model fit, including AIC, BIC, and adjusted Rho-square values. These results imply that a single length of influence zone is not suitable to describe driver behaviour in both homogeneous and HD traffic streams. In this context, one can make the following observations by comparing the parameter estimates of the models for *homogeneous traffic* stream datasets across the three influence zones (Table 4.2 and Table C.1 from Appendix B). First, the influence of the traffic environment variables with respect to MF2 and LF1 and that of the lateral gap between MF1 and LF1 became less influential as the length of the influence zone increased from 30 m to 60 m. This can be observed from the decreasing size of t-statistic values of the corresponding variables as the influence zone increased from 30 m to 60 m. Second, the influence of some variables was not even marginally significant in the later models. For example, the influence of the space gap with respect to LF1 was not significant even at a 70 % confidence level in models with 45 m and 60 m influence zones.

Interestingly, the reverse trend may be observed from similar comparisons of the models estimated on HD traffic stream datasets (Table 4.2 and Table C.2 from Appendix B). That is, the t-statistic values of some of the variables with respect to MF2, LF1, and RF1 increased as the influence zone size increased from 30 m to 60 m. For example, the influence of space gap variables with respect to MF2 and RF1 increased with an increase in the length of the influence zone from 30 m to 60 m. Further, some variables (e.g., space gap with respect to MF2 in the equation for the extent of acceleration) that were found to have an insignificant effect in the model with a 30 m influence zone showed a significant effect in the models with 45 m and 60 m influence zones. That is, drivers in an HD traffic stream consider stimuli from the second lead vehicle (MF2) even if it is as far as 60 m in their decision-making. On the contrary, as discussed earlier, drivers in a *homogeneous traffic* stream consider stimuli from the second lead vehicle if it is closer to their vehicle.

The above differences between homogeneous and HD traffic streams may be attributed to differences in driver behaviour due to differences in the characteristics of the two traffic streams. First, in HD traffic streams, vehicles show a greater extent of lateral movements and can cut in anytime in front of the subject vehicle. Moreover, drivers of influencing vehicles while competing for gaps in the lateral direction might undergo a cycle of deceleration and acceleration. Thus, subject vehicle drivers in HD traffic streams anticipate such behaviour of lead vehicles and consider it in their decision-making even if the lead vehicles are far. Second, drivers in HD traffic conditions do not maintain lane discipline (one can observe vehicles in between lanes, subject vehicles in between multiple leaders – a type of staggered car-following), and the traffic stream itself comprises a higher percentage of motorcycles and three-wheelers which are relatively smaller in size as compared to passenger cars. As a result, the number of lead vehicles within the driver's field of view is higher, and the distance up to which a driver can see (and consider) is larger in HD traffic streams; thus, allowing the driver to perceive stimuli from lead vehicles that are as far as 60 m. On the other hand, drivers in homogeneous conditions either cannot see or may not consider vehicles that are far ahead because drivers adhere to lane discipline (with another vehicle straight ahead), and the traffic stream is dominated by passenger cars along with a small percentage of large vehicles such as trucks.

Table 4.2 Estimation results of the joint models on both homogeneous and HD traffic datasets

Explanatory variables	**Model on homogeneous traffic conditions (influence zone length = 30 m)**				**Model on heterogeneous traffic conditions (influence zone length = 60 m)**			
	Discrete choice#		**Continuous choice***		**Discrete choice#**		**Continuous choice***	
	Accn	**Dccn**	**Accn**	**Dccn**	**Accn**	**Dccn**	**Accn**	**Dccn**
Constant	-0.015 (-0.10)	-0.228 (-1.54)	0.883 (16.55)	0.998 (13.50)	1.992 (5.27)	-0.946 (-2.45)	1.121 (11.94)	-0.067 (-0.69)
Subject vehicle's speed in longitudinal direction (m/s)	0.019 (1.26)	0.080 (6.18)	--	0.021 (4.29)	-0.144 (-4.36)	0.210 (7.39)	-0.025 (-2.86)	0.078 (12.11)
Space gap in longitudinal direction w.r.t. MF1 (m)	0.017 (2.76)	-0.017 (-2.89)	0.006 (2.94)	-0.003 (-1.24)	0.021 (2.97)	-0.017 (-2.52)	0.003 (1.60)	-0.004 (-2.62)
Relative speed in longitudinal direction w.r.t. MF1 (m/s)	0.415 (14.51)	-0.381 (-13.67)	0.068 (5.85)	-0.060 (-4.85)	0.077 (4.08)	-0.081 (-4.12)	0.025 (5.19)	-0.004 (-0.86)
Subject vehicle has 2 or more lead vehicles in MF compartment**	-0.204 (-3.12)	--	--	0.466 (4.34)	--	0.332 (3.84)	--	--
Space gap in longitudinal direction w.r.t. MF2 (m)	--	--	--	-0.021 (-5.04)	0.012 (5.15)	--	0.002 (2.63)	--
Relative speed in longitudinal direction w.r.t. MF2 (m/s)	0.153 (4.35)	-0.263 (-6.86)	0.024 (1.96)	-0.093 (-7.97)	0.082 (6.21)	--	0.014 (2.82)	--
Subject vehicle has 1 or more lead vehicles in LF compartment***	--	--	--	--	--	0.258 (1.72)	--	--
Space gap in longitudinal direction w.r.t. LF1 (m)	--	--	0.002 (1.14)	--	--	-0.008 (-2.57)	0.002 (1.71)	--
Lateral gap between MF1 and LF1 (m)	0.027 (1.28)	--	--	-0.012 (-1.36)	0.106 (4.58)	--	0.009 (1.08)	--
Relative speed in longitudinal direction w.r.t. LF1 (m/s)	--	-0.011 (-1.56)	--	--	0.054 (3.05)	-0.017 (-1.00)	--	--
Subject vehicle has 1 or more lead vehicles in RF compartment***	--	--	--	--	--	0.109 (0.98)	-0.083 (-2.62)	0.096 (3.84)
Space gap in longitudinal direction w.r.t. RF1 (m)	--	--	--	--	--	-0.007 (-2.36)	0.005 (4.27)	--
Lateral gap between MF1 and RF1 (m)	--	--	--	--	--	--	--	--
Relative speed in longitudinal direction w.r.t. RF1 (m/s)	0.014 (1.24)	--	--	--	--	-0.048 (-4.10)	0.019 (4.22)	--
Subject vehicle has 1 or more side vehicle in LS compartment***	--	--	--	--	--	--	--	--
Space gap in lateral direction w.r.t. LS1 (m)	--	--	--	--	0.055 (2.71)	--	--	--
Relative speed in longitudinal direction w.r.t. LS1 (m/s)	--	--	--	--	--	-0.034 (-2.81)	--	--
Subject vehicle has 1 or more side vehicle in RS compartment***	--	--	--	--	-0.283 (-2.26)	--	--	--
Space gap in lateral direction w.r.t. RS1 (m)	--	--	--	--	0.063 (0.97)	--	--	--
Relative speed in longitudinal direction w.r.t. RS1 (m/s)	--	--	--	--	0.048 (2.49)	--	--	-0.019 (-3.23)
Space gap between left edge of the SV and left edge of the road (m)	--	0.013 (2.25)	--	0.002 (1.07)	--	-0.108 (-3.95)	--	--
Scale parameter of regression equations $(\sigma_{\eta i})$			0.625 (45.24)	0.725 (55.10)			0.603 (49.81)	0.577 (57.03)
Scale parameters for panel effects $(\sigma_{\psi i}$ and $\sigma_{\xi i})$								
Acceleration	0.212 (3.15)		--		0.568 (8.43)		--	
Deceleration	0.181 (2.71)		--		0.357 (3.90)		--	
Maintain same speed	0.113 (0.94)		NA		--		NA	
Scale parameters of SV- or driver-level common error terms $(\sigma_{\omega i})$								
Acceleration	0.067 (2.87)				0.144 (7.56)			
Deceleration	--				0.191 (12.62)			
Copula dependency parameters (θ_i)								
Acceleration	-2.896 (-6.61)				-5.257 (-11.05)			
Deceleration	-4.240 (-11.00)				-5.145 (-12.32)			
Goodness of fit measures								
Number of parameters	37				50			
Log likelihood	-14554.46				-10768.14			
AIC value	29182.93				21636.28			
BIC value	29444.68				21978.35			
LLR value w.r.t. to independent model	453.51				878.08			
Critical chi-square value at 95% CI	11.07				9.49			
Adjusted rho-square	0.105				0.143			
Number of vehicles	522				760			
Number of cases	8728				6914			

Notes: Accn = Acceleration, Dccn = Deceleration, # Maintain same speed is base category, *Dependent variable = absolute value of acceleration/deceleration at t s (m/s^2), ** One lead vehicle is base, *** No vehicle is base, -- the corresponding parameter was dropped from the specification as it was statistically insignificant, NA= Not applicable

4.4.3.2 Differences in the number of vehicles influencing the subject vehicle

Another important difference is that the number of vehicles influencing the driver behaviour is greater in HD traffic conditions than in *homogeneous traffic* conditions. Specifically, all six surrounding vehicles have some or the other influence on driver decisions in HD traffic conditions. However, only four out of six vehicles – MF1, MF2, LF1, and RF1 – show an influence on driver behaviour in homogeneous conditions. Notably, the vehicles on the side – LS1 and RS1 – show an influence on driver behaviour in HD traffic conditions only. This may be because, contrary to *homogeneous traffic* conditions, drivers continuously look for gaps in the lateral direction to gain more speed. Another reason is that side vehicles in HD traffic conditions tend to accelerate and cut in front of the subject vehicle. Clearly, drivers in HD traffic conditions have to consider multiple sources of stimuli, including those from the sides. Such lateral interactions are less likely in *homogeneous traffic* conditions unless lane-changing is under consideration.

4.4.3.3 Other differences

In HD traffic conditions, the SV's longitudinal speed shows a significant influence on the decisions to accelerate or decelerate as well as the extent of acceleration and deceleration. The parameter signs indicate that SVs moving at higher speeds are more (less) likely to decelerate (accelerate) and exhibit a higher (lower) magnitude of deceleration (acceleration). In contrast, in *homogeneous traffic* conditions, this variable appears with a positive sign, albeit with a small t-statistic, in the acceleration utility function. This finding seems counterintuitive. However, interestingly, previous studies have also reported similar findings with the SV's speed variable in *homogeneous traffic* conditions (Ahmed, 1999; Toledo, 2003). This may be because the acceleration capabilities of vehicles are higher at high speeds observed in *homogeneous traffic* streams (Toledo, 2003). Another difference between the two traffic conditions is that the SV speed variable does not have a statistically significant influence on the extent of acceleration in *homogeneous traffic* conditions (whereas this variable significantly influences the extent of acceleration in HD traffic conditions). Recall that the influence of subject vehicle speed on the discrete choice utility function for acceleration decision is not statistically significant in the *homogeneous traffic* dataset. It is likely that this variable does not have an influence on the extent of acceleration as well (in the *homogeneous traffic* dataset). It may be that, after accounting for relative speed and headway spacing with respect to the first lead vehicle, the subject vehicle's speed does not have a significant influence on the extent of acceleration in the *homogeneous traffic* dataset. It may also be because we are allowing for the influence of different factors to be different on the discrete and continuous choices. Previous studies that used trajectory datasets from *homogeneous traffic* conditions also report similar findings of the weak influence of subject vehicle speed on its decision to accelerate and/or the extent of acceleration. For example, Ahmed (1999) and Toledo (2003) found that the SV speed did not have a significant influence on the extent of deceleration in *homogeneous traffic* streams on Interstate 93 in downtown Boston and on southbound I-395 in Arlington VA, respectively. The specific reasons behind such findings are not directly apparent and should be explored in future research.

Similarly, after accounting for the aforementioned dependencies and panel effects, the influence of the space gap variable (with respect to MF1) on the extent of deceleration became weak in *homogeneous traffic* streams. Such a weakening of the influence of the space gap variable in the extent of acceleration was observed in HD traffic stream data also, but the final parameter estimates are still statistically significant in the HD traffic conditions. The impact of relative speed with respect to MF1 is also weakened in the extent of deceleration in HD traffic streams after accounting for the dependencies and panel effects.

Note that the gap between the subject vehicle and the left edge of the road is also found to significantly influence driving decisions. A larger gap between the vehicle and the left edge of the road makes car drivers more likely to decelerate (and decelerate more) in *homogeneous traffic* streams. On the contrary, a larger gap makes car drivers tend to decelerate less in HD traffic conditions. This is intuitively aligned with our expectations because it is anticipated that vehicles closer to the left edge drive slowly in India compared to those that are away from the left edge, whereas in the USA, vehicles closer to the right edge drive slowly.

4.5 CONCLUSIONS

In this chapter, we propose a panel data-based discrete-continuous modelling framework to analyse driver behaviour in two disparate trajectory datasets – one from a *heterogeneous, disorderly* (HD) traffic stream in India and another from a *homogeneous traffic* stream in the United States. The panel data-based framework allows the analyst to isolate the subject vehicle- and driver-specific unobserved factors (such as age and aggressiveness) that do not vary across different observations of a same vehicle but have an influence on the vehicle's driver behaviour. Doing so helps reduce the confounding effects of such unobserved factors on analysing the influence of observed factors (such as relative speeds and spacing between the subject vehicle and other vehicles) on driver behaviour. This also helps in reducing the confounding effects of unobserved factors on analysing the differences in driving behaviour between HD and *homogeneous traffic* streams. Furthermore, this chapter employs a simple empirical strategy to remove serial correlation that may arise due to correlation across successive observations from the trajectory of a vehicle.

The estimation results reveal that it is necessary to incorporate the influence of vehicle- and driver-specific unobserved factors while analysing driver behaviour in order to improve the realism of the driver behaviour model. The results revealed interesting similarities and important differences in driver behaviour between homogeneous and heterogeneous traffic conditions trajectory data. From the standpoint of statistical model specification, both homogeneous and heterogeneous datasets revealed that the model that considered the role of unobserved factors at the driver level (*i.e.,* panel data specification) and at the observation level, including dependencies between the discrete and continuous choice equations, provided the best fit. Ignoring such unobserved effects resulted in inflated t-statistics of several parameter estimates and inferior model fit. Since most trajectory datasets pose a panel data setting and dependencies between the discrete and continuous choices of drivers, the choice of the statistical specification becomes an important consideration for estimating the model parameters.

In the context of behavioural similarities, in addition to lead vehicles in the same lane as that of the subject vehicle, vehicles in the lanes adjacent to the subject vehicle (i.e., vehicles in the left front and the right front zones) were also found to influence driver behaviour in both traffic conditions. However, as expected, the influence of the first lead vehicle in the middle front zone (MF1) was found to be the strongest in both traffic conditions, followed by the second lead vehicle (MF2) in the same lane. Further, in both datasets, not all factors that influenced the discrete decisions influenced the continuous decisions and vice versa.

From the standpoint of differences in driver behaviour, the results revealed that side vehicles influence driver decision-making in HD traffic conditions but not in *homogeneous traffic* conditions. Moreover, we found that influence zones of lengths 60 m and 30 m provided the best fit for HD and *homogeneous traffic* datasets, respectively, indicating that drivers in HD traffic streams consider stimuli from lead vehicles even if they are as far as 60 m, whereas drivers in *homogeneous traffic* streams consider stimuli from lead vehicles if they are closer to their vehicle.

The insights from this chapter can assist in developing behaviourally realistic driver behaviour models specific to homogeneous and HD traffic conditions. Specifically, models developed for *homogeneous traffic* conditions might benefit from considering the influence of not only the first lead vehicle but also that of the second lead vehicle in the same lane and that of vehicles in the left front and right front zones within a 30 m influence zone. On the other hand, models for HD traffic in urban settings might benefit from considering the influence of side vehicles (left side and ride side) and that of lead vehicles as far as 60 m ahead of the subject vehicles. Such considerations and the modelling framework presented in this chapter can potentially help in better simulating traffic flow in the two traffic conditions.

CHAPTER 5 DISCRETE CHOICE MODELS WITH MULTIPLICATIVE STOCHASTICITY IN CHOICE ENVIRONMENT VARIABLES: APPLICATION TO ACCOMMODATING PERCEPTION ERRORS IN DRIVER BEHAVIOUR MODELS

Abstract

This chapter 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 as well as 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 on the extent of variability due to perception errors in different traffic environment variables.

Note: The material in this chapter is drawn from the following paper:

Nirmale, S. K., and Pinjari, A. R. (2022). Discrete choice models with multiplicative stochasticity in choice environment variables: application to accommodating perception errors in driver behaviour models. *(In review with Transportation Research Part B: Methodological).*

5.1 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 (Steimetz 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.

5.1.1 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 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, Steimetz and Brownstone (2005) use an imputation method (Rubin, 1987) to correct for measurement errors in network data such as travel times when accurate measurements are available for 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 measurement 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 the 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 the 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.

5.1.2 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.[7]

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 value of time estimation by Hess et al. (2017) motivates the use of multiplicative errors for the utility function to capture context effects, such as a 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

[7] 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 5.2.3, 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.

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 5.2.2, 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 5.2.2, 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 perception errors of the travellers.

5.1.3 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 sources of stochasticity may be due to various reasons – decision-makers' errors in perceiving the choice environment, analyst's error in measuring such variables, or inherent stochasticity of the variables. In this chapter, we focus on the decision-makers' errors in perception as the primary source of stochasticity. The model structure takes the form 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 error 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 a 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 on the magnitudes of variability due to perception errors in different traffic environment variables.

In the rest of this chapter, Section 5.2 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 5.3, details of the vehicle trajectory dataset used in this chapter are presented. Section 5.4 presents the simulation experiments and findings from the experiments. Section 5.5 presents the empirical analysis and discusses the empirical findings. Finally, Section 5.6 summarizes the chapter.

5.2 METHODOLOGY

5.2.1 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 of 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 analyst's measurements (x_{q}) of those variables and the driver behaviour. In this context, it is assumed that the measurements (x_{q}), which are typically obtained from observed vehicle trajectory datasets, are free of errors (see Footnote 7).

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

$$U_{qi} = \beta_{i0} + \sum_{k=1}^{K} \beta_{ik} x_{qk}^* + \xi_{qi} \tag{5.1}$$

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:

$$L_{qi}\left(\beta, x_q^*\right) = \frac{\exp\left(\beta_{i0} + \sum_{k=1}^{K} \beta_{ik} x_{qk}^*\right)}{\sum_{j=a,d,s} \exp\left(\beta_{j0} + \sum_{k=1}^{K} \beta_{jk} x_{qk}^*\right)} \tag{5.2}$$

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:

$$L_{qi}(\beta,\theta) = \int_{x_q^*} L_{qi}\left(\beta, x_q^*\right) f\left(x_q^*;\theta\right) dx_q^* \tag{5.3}$$

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 D provides details on 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. (5.3) 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 5.2.4. Despite such confounding, simulation experiments in Section 5.4 help us identify when a model with stochastic choice environment variables is more suitable than a correlated random coefficients model.

5.2.2 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:

$$x_{qk}^* = x_{qk} + \eta_{qk} \tag{5.4}$$

where, η_{qk} is a normally distributed error component with expected value 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)[8]. 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 5.2.3.

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:

[8] 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.

$$x_{qk}^{*} = x_{qk} . \tau_{qk} \tag{5.5}$$[9]

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 to one, i.e. $E[\tau_{qk}]=1$. This normalization helps in identification as well. In this chapter, 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.

9 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.

5.2.3 *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:

$$\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}^{*}) + \beta_{a2}(x_{q2}^{*}) + \beta_{a3}(x_{q3}^{*}) + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}^{*}) + \beta_{d2}(x_{q2}^{*}) + \beta_{d3}(x_{q3}^{*}) + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned} \tag{5.6}$$

5.2.3.1 Identification for Additive Specification of Error in Choice Environment Variables

Employing the additive error specification of Eq. (5.4) for choice environment variables, the utility structure in Eq. (5.6) may be written as:

$$\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}x_{q1} + \beta_{a2}x_{q2} + \beta_{a3}x_{q3} + \beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}x_{q1} + \beta_{d2}x_{q2} + \beta_{d3}x_{q3} + \beta_{d1}\eta_{q1} + \beta_{d2}\eta_{q2} + \beta_{d3}\eta_{q3} + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned} \tag{5.7}$$

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

$$\begin{aligned}
\varepsilon_{qa} &= \beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa}, \\
\varepsilon_{qd} &= \beta_{d1}\eta_{q1} + \beta_{d2}\eta_{q2} + \beta_{d3}\eta_{q3} + \xi_{qd}, \text{ and} \\
\varepsilon_{qs} &= \xi_{qs}
\end{aligned} \tag{5.8}$$

Without loss of generality, assume that: (a) the additive perception error terms η_{qk} are normally distributed with zero mean and variance $\sigma_{\eta_k}^2$, 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 $\sigma_{\xi j}^2 = \pi^2 / 6g_{\xi}^2$. The variance-covariance matrix of the random utility terms $(\varepsilon_{qa}, \varepsilon_{qd}, \varepsilon_{qs})$ may be derived as below (the derivation of this matrix is provided in Appendix D):

$$\Omega = \begin{bmatrix}
\beta_{a1}^2\sigma_{\eta_1}^2 + \beta_{a2}^2\sigma_{\eta_2}^2 + \beta_{a3}^2\sigma_{\eta_3}^2 + \sigma_{\xi}^2 & \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 & 0 \\
\beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 & \beta_{d1}^2\sigma_{\eta_1}^2 + \beta_{d2}^2\sigma_{\eta_2}^2 + \beta_{d3}^2\sigma_{\eta_3}^2 + \sigma_{\xi}^2 & 0 \\
0 & 0 & \sigma_{\xi}^2
\end{bmatrix} \tag{5.9}$$

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

$$\Omega_{\Delta} = \begin{bmatrix} \beta_{a1}^2\sigma_{\eta_1}^2 + \beta_{a2}^2\sigma_{\eta_2}^2 + \beta_{a3}^2\sigma_{\eta_3}^2 + 2\sigma_{\xi}^2 & \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 + \sigma_{\xi}^2 \\ \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 + \sigma_{\xi}^2 & \beta_{d1}^2\sigma_{\eta_1}^2 + \beta_{d2}^2\sigma_{\eta_2}^2 + \beta_{d3}^2\sigma_{\eta_3}^2 + 2\sigma_{\xi}^2 \end{bmatrix} \quad (5.10)$$

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, which is 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 (non-alternative 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:

$$M_{Rank} = rank[\,jacobian[vecu(\Omega_{\Delta})]] - 1 \quad (5.11)$$

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. (5.10) are:

$$vecu(\Omega_{\Delta}) = \begin{bmatrix} \beta_{a1}^2\sigma_{\eta_1}^2 + \beta_{a2}^2\sigma_{\eta_2}^2 + \beta_{a3}^2\sigma_{\eta_3}^2 + 2\sigma_{\xi}^2 \\ \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2 + \sigma_{\xi}^2 \\ \beta_{d1}^2\sigma_{\eta_1}^2 + \beta_{d2}^2\sigma_{\eta_2}^2 + \beta_{d3}^2\sigma_{\eta_3}^2 + 2\sigma_{\xi}^2 \end{bmatrix} \quad (5.12)$$

and $Jacobian[vecu(\Omega_{\Delta})] = \begin{bmatrix} \beta_{a1}^2 & \beta_{a2}^2 & \beta_{a3}^2 & 2 \\ \beta_{a1}\beta_{d1} & \beta_{a2}\beta_{d2} & \beta_{a3}\beta_{d3} & 1 \\ \beta_{d1}^2 & \beta_{d2}^2 & \beta_{d3}^2 & 2 \end{bmatrix}$. (5.13)

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 number of 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.

5.2.3.2 Identification for Multiplicative Specification of Errors in Choice Environment Variables

Employing the multiplicative error specification of Eq. (5.5) for choice environment variables, the utility structure in Eq. (5.6) may be written as:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}\tau_{q1}) + \beta_{d2}(x_{q2}\tau_{q2}) + \beta_{d3}(x_{q3}\tau_{q3}) + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{5.14}$$

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

$$\begin{aligned} \varsigma_{qa} &= \beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa} \\ \varsigma_{qd} &= \beta_{d1}(x_{q1}\tau_{q1}) + \beta_{d2}(x_{q2}\tau_{q2}) + \beta_{d3}(x_{q3}\tau_{q3}) + \xi_{qd} \\ \varsigma_{qs} &= \xi_{qs} \end{aligned} \tag{5.15}$$

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 domain 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:

$$\Omega = \begin{bmatrix} Var(\varsigma_{qa}) & Cov(\varsigma_{qa},\varsigma_{qd}) & Cov(\varsigma_{qa},\varsigma_{qs}) \\ Cov(\varsigma_{qa},\varsigma_{qd}) & Var(\varsigma_{qd}) & Cov(\varsigma_{qd},\varsigma_{qs}) \\ Cov(\varsigma_{qa},\varsigma_{qs}) & Cov(\varsigma_{qd},\varsigma_{qs}) & Var(\varsigma_{qs}) \end{bmatrix} \tag{5.16}$$

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:

$$\Omega_{\Delta} = \begin{bmatrix} Var(\varsigma_{qa}) + Var(\varsigma_{qs}) - 2Cov(\varsigma_{qa},\varsigma_{qs}) & Cov(\varsigma_{qa},\varsigma_{qd}) + Var(\varsigma_{qs}) \\ Cov(\varsigma_{qa},\varsigma_{qd}) + Var(\varsigma_{qs}) & Var(\varsigma_{qd}) + Var(\varsigma_{qs}) - 2Cov(\varsigma_{qd},\varsigma_{qs}) \end{bmatrix} \tag{5.17}$$

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.

5.2.4 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) 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 5.4 and 5.5, respectively.

5.2.5 Comparison with the Random Coefficients Specification

In the context of the multiplicative stochasticity specification as in Eq. (5.5) 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:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \sum_{k=1}^{K} \beta_{ak} x_{qk} \tau_{qk} + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \sum_{k=1}^{K} \beta_{dk} x_{qk} \tau_{qk} + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{5.18}$$

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 alterative specific and their distributions would be different, even if their standard deviations are of 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.[10]

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

[10] 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.

problems during estimation. To understand this better, consider the following utility functions with correlated random coefficients that are lognormally distributed:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{akRC} - \sigma_{ak}\Phi^{-1}\left[1-\Phi\left(Z_{qak}\right)\right]\right) x_{qk} + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{dkRC} - \sigma_{dk}\Phi^{-1}\left[1-\Phi\left(\rho_k Z_{qak} + \left(\sqrt{1-\rho_k^2}\right) Z_{qdk}\right)\right]\right) x_{qk} + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{5.19}$$

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:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{akRC}\right) \times \exp\left(\sigma_k Z_{qak} - 0.5\sigma_k^2\right) \times x_{qk} + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \sum\nolimits_{k=1}^{K} \exp\left(\beta_{dkRC}\right) \times \exp\left(\sigma_k Z_{qak} - 0.5\sigma_k^2\right) \times x_{qk} + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{5.20}$$

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. (5.18), 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 in Eq. (5.18) 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 5.4.4). In such situations, only if the above-mentioned constraints are imposed, a CRC model would be identified. 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.

5.2.6 Alternative Distributions for Multiplicative Errors in Choice Environment Variables

As indicated earlier, choice environment variables representing physical quantities such as distance, time, and speeds 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 D.2 in Appendix D 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 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.

5.3 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 Chapter 3, we 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 5.1. 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.

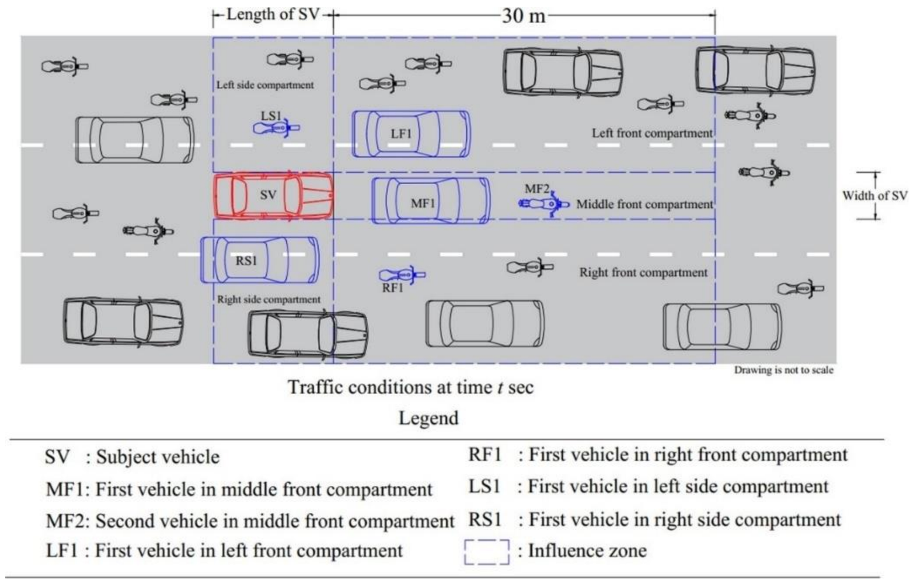


Figure 5.1 Structure of influence zone around a subject vehicle

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 $t-0.5\,\text{s}$ (note: 0.5 s is considered the reaction time, based on an analysis in Chapter 3) (b) all other vehicles and their characteristics (type and dimensions) and infrastructure elements within the influence zone at $t-0.5\,\text{s}$, 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 $t-0.5\,\text{s}$. 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.

5.4 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.

5.4.1 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 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}^{*}), perceived longitudinal space gap between SV and MF1 (x_{q2}^{*}), perceived relative speed of between SV and MF1 (x_{q3}^{*}), and 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:

$$\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \beta_{a4}(x_{q4}\tau_{q4}) + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \beta_{d2}(x_{q2}\tau_{q2}) + \beta_{d3}(x_{q3}\tau_{q3}) + \beta_{d4}(x_{q4}\tau_{q4}) + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned} \tag{5.21}$$

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 5.1 (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.

5.4.2 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) = $\sqrt{(\text{mean estimate - true value})^2 + FSSE^2}$

A summary of the above performance measures is presented in Table 5.1 for each parameter, along with the true value of each parameter. As can be observed from the table, the proposed model was able to accurately recover parameters 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.[11]

Table 5.1 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

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

[11] 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.

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 on which variables are associated with greater variability than others.

5.4.3 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 (5.21), the following alternative ML models were estimated on the same simulated data from Section 5.4.1:

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

$$\begin{aligned} U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \sum_{k=2,3,4} (\tau_{qak}) x_{qk} + \xi_{qa} \\ U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \sum_{k=2,3,4} (\tau_{qdk}) x_{qk} + \xi_{qd} \\ U_{qs} &= \xi_{qs} \end{aligned} \tag{5.22}$$

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. (5.22), 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:

$$\begin{aligned} \tau_{qak} &= \exp\left(\beta_{ak} - \sigma_{k1} \Phi^{-1} \left[\left(1 - \Phi\left(Z_{qk1} \right) \right)^{1/p_{k1}} \right] \right) \\ \tau_{qdk} &= \exp\left(\beta_{dk} - \sigma_{k2} \Phi^{-1} \left[\left(1 - \Phi\left(\rho_k Z_{qk1} + \left(\sqrt{1 - \left(\rho_k \right)^2} \right) Z_{qk2} \right) \right)^{1/p_{k2}} \right] \right) \end{aligned} \tag{5.23}$$

Recall (from Section 5.2.5) that the correlated random coefficients model subsumes the multiplicative EIV model as a special case when the corresponding random coefficients are perfectly correlated (with 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:

$$\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \sum_{k=2,3,4} \beta_{ak} x_{qk} + \tau_{qad} + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \sum_{k=2,3,4} \beta_{dk} x_{qk} + \tau_{qad} + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned} \tag{5.24}$$

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:

$$\begin{aligned}
U_{qa} &= \beta_{a0} + \beta_{a1}(x_{q1}) + \sum_{k=2,3,4} \beta_{ak} x_{qk} + \tau_{qa} + \xi_{qa} \\
U_{qd} &= \beta_{d0} + \beta_{d1}(x_{q1}) + \sum_{k=2,3,4} \beta_{dk} x_{qk} + \tau_{qd} + \xi_{qd} \\
U_{qs} &= \xi_{qs}
\end{aligned} \tag{5.25}$$

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 5.2 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 necessarily help in capturing 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 5.2 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 5.2.5 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.

5.4.4 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 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 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 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 correlation level is low (scale parameters are 0.80 and 1.50, ρ = 0.3).

Table 5.3 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 a majority 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.[12]

[12] 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

Table 5.3 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

5.4.5 Guidance for Model Selection

Based on the conceptual discussions in Section 5.2 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.
- Considering that most empirical research involves moderate-sized datasets of a few thousand samples or less, use the AIC metric to determine a preferred model structure. In addition to the data fit metrics to select a model structure, use the following guidelines for interpretation.
- If the CRC model estimation shows signs of unidentifiability (as discussed in Section 5.4.4) 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

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.

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 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 positive correlation between the acceleration and deceleration choice alternatives. Several unobserved factors such as driver demographics and aggressiveness are likely to be associated with opposite preferences between acceleration and deceleration decisions. So, positive correlation between random coefficients of such choice alternatives, if any, is likely due to drivers' perception errors. 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. In essence, the analyst should combine statistical fit and intuitive considerations to decide the model structure and its interpretation.

5.5 EMPIRICAL ANALYSIS

5.5.1 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 5.2) 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 5.2.6. 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 the 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, a subsequent empirical analysis of the ML-ME model was conducted with PLN distribution for the 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 5.4.3 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 5.4.3, 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. Even for this variable, the estimated correlation parameter between the random coefficients in acceleration and deceleration utility functions was not statistically different from 1. Such perfect correlation suggests stochasticity in the variable (not in coefficients of the variable).

5.5.2 Goodness of Fit in Estimation and Validation Datasets

Table 5.4 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:

$$\text{Significance Level} = \Phi\left[-\left(-2\left(\rho_H^2-\rho_L^2\right)\times LL(0)+\left(K_H-K_L\right)\right)^{\frac{1}{2}}\right] \tag{5.26}$$

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 5.4 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

5.5.3 *Empirical Findings*

Table 5.5 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 reported in Appendix D.

5.5.3.1 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 as well as interpretation among all other specifications.

The bottom set of rows in Table 5.5 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).

Table 5.5 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. Maintain same speed is the base alternative. ** 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.

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 in 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 of 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:

$$Std_{PLN} = \sqrt{\left\{ \int_0^1 \exp\left(-2\sigma\Phi^{-1}\left(y^{1/p}\right) + 2\mu\right) dy \right\} - Mean^2} \tag{5.27}$$

The corresponding standard deviation parameters are reported in Table 5.5 in the column titled "Standard deviation." Comparing the magnitudes of 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 lateral gaps between two moving vehicles is more difficult than perceiving 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.

5.5.3.2 Empirical Findings other than Perception Errors

The empirical model results in Table 5.5 offers interesting insights from the ML-ME-PLN model. 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 deceleration decision is greater than that for 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 Chapter 3, 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.

5.6 CONCLUSIONS

This chapter 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 the 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 a 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 vehicles' (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 (un)identifiability 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 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.

CHAPTER 6 A TWO-DIMENSIONAL, MULTI-VEHICLE ANTICIPATION, AND MULTI-STIMULI BASED LATENT CLASS FRAMEWORK TO MODEL DRIVER BEHAVIOUR IN HETEROGENEOUS, DISORDERLY TRAFFIC CONDITIONS

Abstract

This study formulates a latent class-based driving behaviour framework for modelling vehicles' two-dimensional (2D) movements while considering drivers' strategic intents and multi-vehicle anticipation (MVA) in heterogeneous, disorderly (HD) traffic conditions. Specifically, five extensions are proposed to a typical stimulus-response based driving behaviour framework. First, the subject vehicle's 2D movements are represented as a combination of the angular direction of movement with respect to the longitudinal axis and the magnitude of acceleration or deceleration along the angle. Second, a latent class framework is used to recognize drivers' strategic intents (latent to the analyst) in two dimensions: (a) the intent to accelerate, decelerate, or maintain a constant speed, and (b) the intent to steer to the left of, right of, or straight along the longitudinal axis. It is hypothesized that these strategic intents precede tactical decisions such as how much to accelerate or decelerate and which specific angular direction to move along. Third, the MVA effect is accommodated to recognize that drivers consider stimuli from multiple vehicles in their vicinity. Fourth, a multi-stimuli model of acceleration (deceleration) is formulated assuming that drivers choose an angle of movement that allows them to move with the highest (lowest) possible longitudinal acceleration (deceleration). Fifth, drivers' execution errors are recognized as the difference between their planned acceleration and executed acceleration. The proposed framework is applied for an analysis of motorised two-wheeler driver behaviour using a vehicular trajectory dataset from India. The empirical results highlight the importance of incorporating MVA and considering driver's intents while modelling 2D movements of vehicles in HD traffic conditions. Further, the microscopic traffic environment variables are found to have a stronger influence on drivers' higher-level, strategic intents than on their lower-level, tactical decisions.

Note: The material in this chapter is drawn from the following paper:

Nirmale, S. K., Pinjari, A. R., and Chakroborty P. (2022). A Two-Dimensional, Multi-Vehicle Anticipation, and Multi-Stimuli Based Latent Class Framework to Model Driver Behaviour in Heterogeneous, Disorderly Traffic Conditions. *(In review with Transportation Research Part C: Emerging Technologies).*

6.1 INTRODUCTION

Single-leader car-following models, where a subject vehicle (SV) is assumed to follow a single lead vehicle (LV) ahead of it, have been extensively used to model driver behaviour in both homogeneous and heterogeneous, disorderly (HD) traffic conditions. Further, most literature in this area focuses on vehicles' longitudinal movement, while lane-changing models are used to consider lateral movements separately from longitudinal movements. However, emerging literature (Kanagaraj and Treiber, 2018; Sarkar et al., 2020) and some early studies (Bexelius, 1968; Lenz et al., 1999; Hoogendoorn and Ossen, 2006) argue that assuming a single leader's influence and/or separating longitudinal movements from lateral movements may not adequately represent driver behaviour. There is increasing evidence of multi-vehicle anticipation (MVA), where drivers consider influences (or stimuli such as space gaps and relative speeds) from multiple vehicles in the vicinity of their vehicle to make their manoeuvring decisions. Recent studies suggest a greater extent of the MVA effect in HD traffic conditions observed in countries such as India than that in homogeneous traffic conditions. Further, unlike in homogeneous traffic streams, in HD traffic streams, drivers' manoeuvres are not only restricted to longitudinal vehicle-following. Instead, a considerable portion of vehicular movements tends to be two-dimensional (2D), where vehicles might move in an oblique direction that results in a simultaneous movement along the longitudinal and lateral directions. The extent of such 2D movement has been observed to be substantial for motorised two-wheelers, which comprise a significant portion of the traffic mix in HD traffic streams. Therefore, driver behaviour in HD traffic conditions is better represented by considering both 2D movements and the MVA effect. Only a few studies (Lee et al., 2009; Shiomi et al., 2012; Amrutsamanvar, 2020; Sarkar et al., 2020) consider the influence of multiple vehicles while modelling 2D movements of a subject vehicle.

Furthermore, the driving actions observed in vehicle trajectory datasets typically available to the analysts do not reveal the drivers' actual intents. They only reveal the final driving actions taken on the road, in the form of the extent of acceleration or deceleration and the angular direction of the movement with respect to the direction of traffic flow (i.e., the longitudinal direction). These driving actions or manoeuvres executed by the drivers while travelling on a road are typically preceded by their intents to accelerate, decelerate, or maintain a constant speed and to steer to the left of, right of, or straight along the longitudinal direction. However, very small acceleration or deceleration values and slight angular deviations from the direction of traffic flow are common even if the drivers intend to maintain a constant speed state and to move in a straight path, respectively. This is due to the difficulty in maintaining zero acceleration and avoiding any lateral movement. Therefore, it is helpful to consider drivers' intents while modelling their 2D movements. However, as mentioned earlier, the actual intents of drivers are unobserved or latent to the analyst, and only the outcome of the driver's actions (such as the extent of acceleration and angular deviation from the direction of traffic flow) can be observed in the vehicle trajectory datasets. While some literature exists on modelling drivers' intents or latent plans in homogeneous traffic streams (Choudhury, 2007; Koutsopoulos and Farah, 2012; Choudhury and Islam, 2016), none exists on modelling driver intents in HD traffic streams while considering 2D movements.

As discussed above, driving manoeuvres in HD traffic streams involve multifaceted decisions such as: (a) the intent to accelerate, decelerate, or maintain a constant speed, (b) the extent of acceleration or deceleration, (c) the intent to steer or orient to the left of, right of, or straight along the traffic flow direction, and (d) the specific angular direction of movement. These decisions must be made quickly based on the drivers' perceptions of the constantly evolving traffic environment around them. However, from a cognitive science standpoint, humans are endowed with a limited amount of cognitive resources such as working memory they need to store and process information for making decisions (Sweller, 1988). Therefore, drivers might allocate their cognitive resources optimally to quickly make their manoeuvring decisions. Specifically, given the multifaceted decisions drivers need to make in a short timeframe and the complexity of the traffic environment around them, it is plausible that they break down their decision-making into manageable steps for cognitive ease. Limited efforts have been made in the literature to explore behavioural mechanisms of how drivers might break down complex driving tasks. In this context, there is scope to explore if analysing drivers' intents (which are latent to the analyst) first, followed by the specific actions they take, can help in filling this gap (see Choudhury, 2007; Koutsopoulos and Farah, 2012; Choudhury and Islam, 2016 for some work in this direction). For instance, it may be that higher-level, strategic decisions – such as the intents of whether to accelerate, decelerate, or maintain a constant speed and whether to steer to the left of, right of, or keep straight along the longitudinal direction – are made first, followed by lower-level, tactical decisions – such as exactly how much to accelerate or decelerate and which specific angular direction to move along. In this context, we conjecture that a greater amount of cognitive effort might be invested in making the higher-level, strategic decisions or intents than that for making lower-level, tactical decisions. This is likely because, as will be seen in the empirical analysis, a greater amount of information is processed for making higher-level intents than that for making lower-level decisions. The current study aims at gathering evidence toward this conjecture using statistical analysis of the trajectory data, without necessarily delving into measurements of the cognitive loads exerted in making the above-mentioned decisions.

In the existing literature, the latent class modelling approach has been used to consider latent (to the analyst) intents or plans of the drivers for modelling driving behaviour (Choudhury, 2007; Koutsopoulos and Farah, 2012; Choudhury and Islam, 2016). In another stream of literature, various approaches have been employed to model drivers' 2D movement. These include (but are not limited to) the force field approach (Chakroborty et al., 2004; Kanagaraj and Treiber, 2018), the simulation-based approach (Maurya, 2007b; Mathew et al., 2013), the two-dimensional continuum modelling approach (Vikram et al., 2022), and the utility theory-based models (Lee, 2007; Shiomi et al., 2012; Sarkar et al., 2020). A third stream of literature highlights the influence of the MVA effect on driver behaviour. However, most studies in the literature considered the above-discussed aspects – drivers' intents latent to the analyst, 2D movement of vehicles, and the MVA effect – in isolation. Such studies in the context of HD traffic conditions are even more sparce in the literature. Importantly, we are not aware of any driver behaviour model that considers drivers' intents and 2D movements simultaneously while also recognizing the MVA effect for analysing driver behaviour in HD traffic streams. The current study fills this specific gap by formulating a latent class-based driving behaviour framework that considers drivers' intents for modelling vehicles' 2D movements while considering the MVA effect on these movements in HD traffic situations. Specifically, the following extensions are proposed to a conventional stimulus-response based driving behaviour framework:

1. The observed 2D movements of a subject vehicle (SV) are represented as a combination of: (a) the angular direction (or orientation) of movement with respect to the longitudinal direction and (b) the magnitude of acceleration or deceleration along the direction.
2. The observed 2D movements of an SV at a time instance are assumed to be a result of a sequential decision-making process of its driver, where higher-level (strategic) decisions precede lower-level (tactical) decisions. The higher-level decisions are drivers' intents along the following two dimensions:
    (a) the intent to accelerate, decelerate, or maintain a constant speed, and
    (b) the intent to steer to the left of, right of, or straight along the longitudinal direction.

The lower-level decisions are the tactical decisions of how much to accelerate or decelerate and which specific angular direction to move along.

A latent class framework is utilised to model drivers' intents as they are latent to the analyst. Conditional on the intents, the lower-level decisions are modelled. The sequential decision-making assumption combined with our two-stage modelling framework allows us to gather evidence toward our conjecture that the extent of cognitive effort needed for making higher-level intents is different from those for making lower-level choices. In the absence of measurements of cognitive load, this is achieved by comparing the strength of influence of various traffic environment variables on the higher- and lower-level decisions.

3. The MVA effect is accommodated, wherein the driver of an SV is assumed to consider stimuli from multiple vehicles within its two-dimensional influence zone for making their intents and manoeuvring decisions.
4. For the lower-level decisions, a multi-stimuli model of acceleration is formulated based on the assumption that drivers choose a specific angle of movement that allows them to move with the highest (lowest) possible longitudinal acceleration (deceleration) if they intend to accelerate (decelerate).
5. Drivers' execution errors are modelled as the difference between their planned extent of acceleration or deceleration and the executed acceleration values.

Finally, an empirical application of the proposed framework for motorised two-wheelers is presented using an HD traffic conditions trajectory dataset from Chennai, India.

The rest of this chapter is structured as follows. Section 6.2 describes the proposed modelling framework. Section 6.3 discusses the Chennai trajectory dataset and variables used to build the empirical model. Section 6.4 presents the empirical model results (using trajectory data from Chennai) and discusses insights on the driver behaviour of motorised two-wheelers in HD traffic streams. Finally, Section 6.5 summarises the main findings of this chapter.

6.2 METHODOLOGY

As discussed earlier, the observed extent of acceleration (or deceleration) and the specific steering direction executed by a driver while travelling on the road are preceded by her/his intents. Therefore, the proposed modelling framework analyses the driver's intents and her/his vehicle's 2D movements using a two-stage modelling framework that comprises the following components: (a) a latent class model component for analysing the driver's higher-level decisions or intents, and (b) intent-specific models (class-specific models) for analysing the driver's lower-level decisions such as the specific steering angle and the extent of acceleration or deceleration. In addition, the framework incorporates variability due to execution error – the difference between the driver's planned (or modelled) acceleration values and the executed acceleration values. Each of these model components is discussed next.

6.2.1 Latent Class Model Component for Analysing Higher-Level Decisions (Intents)

Since a driver's higher-level intents cannot be observed in trajectory datasets, they are probabilistically mapped to the observed manoeuvring actions of the driver, as shown in Figure 6.1. In this figure, the driver's intents are shown in ovals, while the observed manoeuvring outcomes are shown in rectangles. This is done separately for two different types of intents – the first type is the *acceleration intent*, which pertains to the intent to accelerate (A), decelerate (D), or maintain a constant speed (C) and the second type is the *steering direction intent*, which pertains to the intent to steer to the left of (L), right of (R), or straight along (S) the longitudinal direction. Note that the observed steering angles to the left of the longitudinal axis are denoted as +ve angles and those to the right of the longitudinal direction are denoted as -ve angles.

The overall set of latent intents considered in this study comprises a combination of acceleration intents and steering direction intents – a total of nine possible intent combinations given by the set $J = \{AL, AS, AR, CL, CS, CR, DL, DS, DR\}$. In the elements of this set, the letters A, C, and D correspond to the acceleration intents and the letters L, S, and R correspond to the steering direction intents. For example, AL corresponds to the driver's intent to accelerate and steer left of the longitudinal direction.

The logit function in Eq. (6.1) is used to model the driver's intents that are latent to the analyst:

$$P_{q(t+\Delta t)}(s) = \frac{\exp\left(\sum_{k=1}^{K} \beta_{sk} x_{qtk}\right)}{\sum_{s \in J} \exp\left(\sum_{k=1}^{K} \beta_{sk} x_{qtk}\right)}, \; J = \{AL, AS, AR, CL, CS, CR, DL, DS, DR\} \qquad (6.1)$$

Here, $P_{q(t+\Delta t)}(s)$ is the probability that the driver of a subject vehicle q at time $(t+\Delta t)$ intends to make a higher-level decision (or intent) s from the set J of the possible intents; x_{qtk} represents the k^{th} explanatory variable in the set of K traffic environment variables (e.g., space gaps and relative speeds with respect to surrounding vehicles at time t) influencing the driver's intent; and β_{sk} is the corresponding parameter to be estimated. Note that Δt is the reaction time assumed to be constant for all drivers (in future research, it would be useful to relax this assumption to allow heterogeneity in reaction times (Toledo, 2003)).

It is easy to consider the MVA effect in the above expression by including stimuli (space gaps and relative speeds) from multiple surrounding vehicles as explanatory variables. Note here that the effect of traffic environment variables need not be specified separately in each intent s from the set J. Instead, the analyst can employ a parsimonious, dimension-wise specification (more on this later).

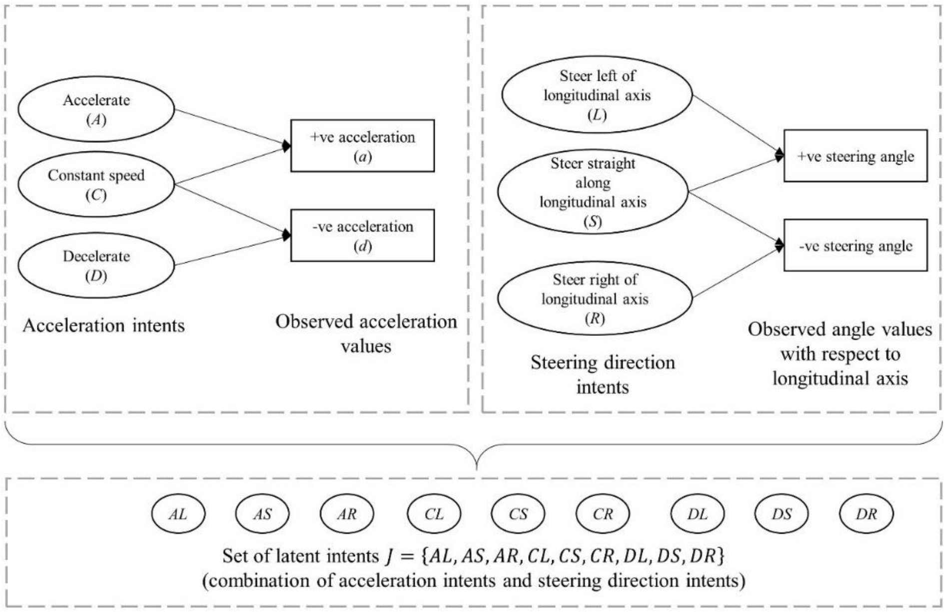


Figure 6.1 Defining latent intents based on observed manoeuvring actions of drivers

6.2.2 *Intent-Specific Model Components for Analysing Lower-Level Decisions*

In this study, conditional on the higher-level decisions (or intents), drivers are assumed to make the following lower-level decisions: (a) a specific steering angle and (b) the specific extent of acceleration (if the intent is to accelerate) or deceleration (if the intent is to decelerate). Since there are nine possible higher-level decisions (intents), we specify the models of steering angle choice and the extent of acceleration or deceleration for each of these intents. To do so, some notational preliminaries are provided next.

To model the steering angle, we discretise the two-dimensional space ahead of the subject vehicle (SV) into a set of six angles, as shown in Figure 6.2. Each of these angles is represented by a ray i beginning from the subject vehicle (vertex of the angle) and bisecting the angle. In the set I of these rays, where $I = \{l_1, l_2, l_3, r_1, r_2, r_3\}$, the rays l_1, l_2, l_3 represent angles 0^o to 1^o, 1^o to 3^o, and $> 3^o$, respectively, in the anticlockwise direction. And the rays r_1, r_2, r_3 represent angles 0^o to 1^o, 1^o to 3^o, and $> 3^o$, respectively, in the clockwise direction. For each of these angles, the angle measure θ_i extended between the longitudinal axis and the corresponding ray i is taken as it's steering angle value. For any vehicle steering along a given angular direction i, it is assumed to be steering at angle θ_i with respect to the longitudinal axis. The set of $\theta_i \, \forall i \in I$ corresponding to all rays/angular directions in the set I is $\Theta = \{0.5^o, 2^0, 9^o, -0.5^o, -2^0, -9^o\}$. In this set, 9^o and -9^o are taken as angle values for l_3 and r_3, respectively, as observed data did not show many vehicles steering beyond $\pm 15^o$ to the longitudinal axis.

Note that the choice set of steering angles (θ_i) considered by a driver depends on their steering angle intent (refer Figure 6.3). That is, if they intend to steer left of the longitudinal axis, they are assumed to consider the angles represented by $I_L = \{l_1, l_2, l_3\}$ or the steering angle values given by $\Theta_L = \{0.5^o, 2^0, 9^o\}$. If they intend to steer right, they would consider the angles represented by $I_R = \{r_1, r_2, r_3\}$ or the steering angles values given by $\Theta_R = \{-0.5^o, -2^0, -9^o\}$. If they intend to steer straight (i.e., along the longitudinal axis), they would consider the angles represented by $I_S = \{l_1, l_2, r_1, r_2\}$ or the steering angle values given by $\Theta_S = \{0.5^o, 2^0, -0.5^o, -2^0\}$, as no one would intend to keep straight but end up steering beyond $\pm 3^0$ (of the longitudinal axis).

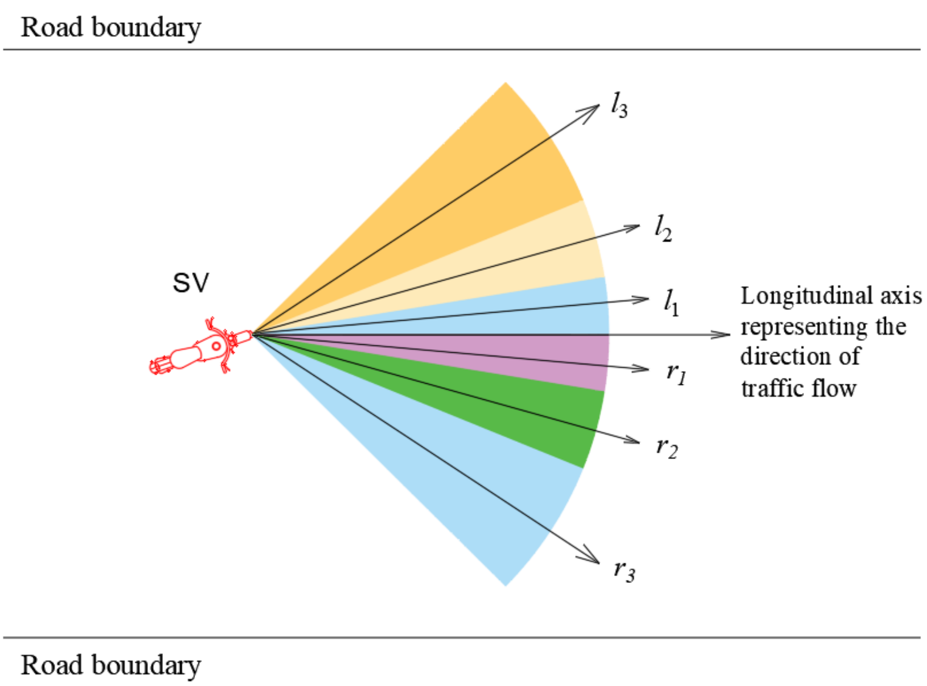


Figure 6.2 Division of the 2D roadway space ahead of a vehicle into steering angles

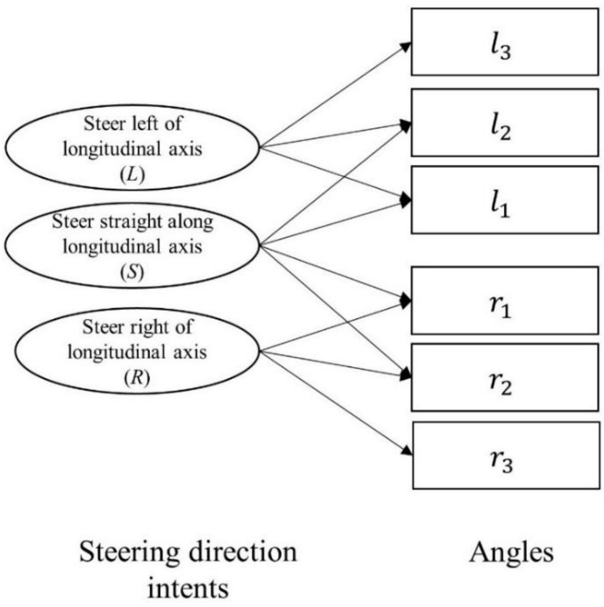


Figure 6.3 Mapping between steering direction intents (s) and steering angles (i)

6.2.2.1 Steering angle choice models conditional on the intents AL, AR, DL, and DR

For the intents AL and AR, the drivers are assumed to choose a specific steering angular direction i (or the corresponding steering angle θ_i) that allows them to proceed with the maximum possible longitudinal acceleration. If the intents are either DL or DR, they are assumed to choose a steering angle that allows them to proceed with the minimum possible longitudinal deceleration. This is because drivers prefer to arrive at their destination as quickly as possible without spending additional time than necessary on the road (Chakroborty et al., 2004). Therefore, the model for the choice of a specific steering angular direction i conditional on intents $AL, AR, DL,$ or DR is specified as an acceleration maximisation model (or deceleration minimisation model) as below:

$$P_{q(t+\Delta t)}\left(i \middle| s \in \{AL, AR, DL, DR\}\right) = P\left(A^s_{q(t+\Delta t)i} \geq A^s_{q(t+\Delta t)j} \forall j \in G^s_{q(t+\Delta t)}\right) \tag{6.2}$$

Here, $A^s_{q(t+\Delta t)i}$ is the specific longitudinal acceleration value with which the driver can proceed along the angular direction i at time $(t+\Delta t)$ conditional on the higher-level intent being s. $G^s_{q(t+\Delta t)}$ is the choice set of angular directions available for the driver conditional on the intent s (see Figure 6.3 and relevant discussion). Note that Eq. (6.2) uses an acceleration maximisation formulation for both acceleration and deceleration intents because the term $A^s_{q(t+\Delta t)i}$ can take both positive values (for acceleration) and negative values (for deceleration).

6.2.2.2 Acceleration model conditional on the intents $AL, AR, DL,$ and DR

We propose the following multi-stimulus response-based model for $A^s_{q(t+\Delta t)i}$ in Eq. (2):

$$\begin{aligned} A^s_{q(t+\Delta t)i} &= \cos(\theta_i) \times f_s\left(AD_{qti}\right) \times \begin{bmatrix} \alpha_{s1} \\ +\alpha_{s2}\left(V^{des} - V_{qt}\right) \\ +\alpha_{s3}\left[\lambda_i\left(\Delta X_{qti} - \Delta X^{des}_{qt}\right) + \left(1-\lambda_i\right)\left(\Delta X^{curb}_{qti} - \Delta X^{des}_{qt}\right)\right] \\ +\alpha_{s4}\lambda_i \Delta V_{qti} \end{bmatrix} + \varepsilon^s_{qti} \\ &= a^s_{q(t+\Delta t)i} + \varepsilon^s_{qti} \end{aligned} \tag{6.3}$$

In the above model, the following three different types of stimuli are considered for the longitudinal acceleration that a subject vehicle (SV) can proceed with along the direction i at time $(t+\Delta t)$:

(1) The difference between the driver's desired longitudinal speed (V^{des}) and the subject vehicle's longitudinal speed (V_{qt}) at time t, assuming that drivers prefer to drive at a given desired speed.

(2) The available space ahead of the SV along the direction i, after subtracting the minimum space gap (ΔX^{des}_{qt}) acceptable to drivers. This space is computed as

$\lambda_i\left(\Delta X_{qti}-\Delta X_{qt}^{des}\right)+\left(1-\lambda_i\right)\left(\Delta X_{qti}^{curb}-\Delta X_{qt}^{des}\right)$, where λ_i is a dummy variable coded as 1 if a vehicle is present ahead of the SV along the direction i (0 otherwise), ΔX_{qti} is the space between the SV and the vehicle ahead of SV along the direction i, and ΔX_{qti}^{curb} is the space between the SV and the curb ahead along the direction i.

(3) the relative velocity (ΔV_{qti}) between the SV and the vehicle ahead of it along the direction i, if there is a vehicle ahead.

The parameters α_{s2}, α_{s3}, and α_{s4} capture the influence of the above three stimuli, respectively, on the extent of acceleration. These intent-specific parameters can be estimated using empirical trajectory data. The variables V_{qt}, ΔX_{qti}, ΔV_{qti}, and λ_i are observed from the empirical data. However, the values of a driver's desired speed and space gap (V^{des} and ΔX_{qt}^{des}) must be specified by the analyst because they cannot be estimated from the trajectory data. In this study, V^{des} is assumed as 60 km/hr for all drivers, based on the speed limit of the roadway stretch from which trajectory data was obtained for the empirical analysis. Next, for ΔX_{qt}^{des}, we employed Pipes (1953)'s modified equation as below:

$$\Delta X_{qt}^{des}=\Delta X_{\min}+\frac{L_q}{0.447\times 10}V_{qt} \tag{6.4}$$

Here, $\Delta X_{\min}$ is the minimum gap drivers prefer to maintain with respect to a vehicle ahead when their vehicle is at rest in jam density situation. We adopt $\Delta X_{\min}$ as the length of the subject vehicle (L_q). The term $\frac{L_q}{0.447\times 10}V_{qt}$ assumes that the SV needs at least L_q additional gap for every additional speed of 10 mph.

Admittedly, the values for V^{des} and $\Delta X_{\min}$ are assumed to be the same for all drivers. However, these quantities may vary across drivers. To capture the average effect of such heterogeneity and that of any additional influences on the SV's acceleration, a constant α_{s1} is incorporated into the acceleration model. This constant can be estimated from the empirical data. Note that changing the values assumed for V^{des} and $\Delta X_{\min}$ does not change the model, in terms of its fit (likelihood) to the empirical data, the interpretation or the statistical significance of other parameters – only the magnitudes of the parameters α_{s1}, α_{s2}, α_{s3}, and α_{s4} change. Nevertheless, it is preferable that the values assumed for V^{des} and $\Delta X_{\min}$ reflect the travel conditions on the road.

Next, the model in Eq. (6.3) for the extent of acceleration (or deceleration) for a steering angle θ_i should recognise that a vehicle may need to change its steering angle (θ_{qt}) at t to the steering angle θ_i at $(t+\Delta t)$. To recognise this, we use $AD_{qti} = |\theta_{qt} - \theta_i|$ to measure the change in steering angle from t to $(t+\Delta t)$. Since it is more difficult to execute a larger change in the steering angle, one can expect that the extent of acceleration achieved in effecting a larger change in the steering angle would be smaller. Therefore, when the intent is to accelerate $(s = AL \text{ or } AR)$, we employ a decreasing function of AD_{qti} to moderate the level of acceleration achieved for executing AD_{qti} amount of change in angular orientation. For the same reason, when the intent is to decelerate $(s = DL \text{ or } DR)$ we employ an increasing function of AD_{qti}. We explored different functional forms for the moderating functions to incorporate the impact of change in steering angle, and found that the following functions provided the best fit to the empirical data:

$$f_s\left(AD_{qti}\right) = \begin{cases} \dfrac{2\exp\left(\dfrac{-AD_{qti}\times\pi}{180}\right)}{1+\exp\left(\dfrac{-AD_{qti}\times\pi}{180}\right)} & \forall \mathrm{s} \in \{AL, AR\} \\[2ex] \dfrac{2\exp\left(\dfrac{AD_{qti}\times\pi}{180}\right)}{1+\exp\left(\dfrac{AD_{qti}\times\pi}{180}\right)} & \forall \mathrm{s} \in \{DL, DR\} \end{cases} \tag{6.5}$$

Finally, the expression in the square bracket of Eq. (6.3), along with its product with $f_s\left(AD_{qti}\right)$, provides the acceleration along the direction i conditional on intent s. To obtain the longitudinal acceleration, the expression is multiplied with $\cos\left(\theta_i\right)$. Further, to account for variation in the accelerations due to inter-driver heterogeneity in V^{des} and $\Delta X_{\min}$ (which are assumed to be same for all drivers), and any other influential factors not considered in the model, a stochastic term ε_{qti}^{s} is introduced in Eq. (6.3). This term is assumed to be normally distributed with zero mean and unit variance.

With the above assumptions, embedding Eq. (6.3) for the extent of acceleration or deceleration into the steering angle choice model for Eq. (6.2) results in a multinomial probit type discrete choice model. Further, the acceleration value conditional on the intent $s \in \{AL, AR, DL, DR\}$ may be expressed as:

$$A^s_{q(t+\Delta t)} = \underset{j \in G^s_{q(t+\Delta t)}}{Max} \left(A^s_{q(t+\Delta t)j} \right) \tag{6.6}$$

6.2.2.3 Steering angle choice and acceleration models conditional on the intents AS and DS

Intuitively, when drivers intend to steer in the straight direction (i.e., $s \in \{AS, DS\}$), they would not think about choosing an "angular direction." At the same time, they might unintendedly drift along a direction that is slightly left or right of the longitudinal axis. Therefore, unlike the acceleration-maximisation-based steering angle choice model in Eq. (6.2), the choice of angular direction i conditional on intents $s \in \{AS, DS\}$ is modelled using a simple discrete choice framework with only constants in the specification. That is, the choice of the specific angle is not coupled with the extent of the acceleration.

To model the extent of acceleration or deceleration $A^s_{q(t+\Delta t)}$ for intents $s \in \{AS, DS\}$, we use the same expression as in Eq. (6.3) for $A^s_{q(t+\Delta t)}$ for the chosen angular direction i.

6.2.2.4 Steering angle choice and acceleration models conditional on the intents $CL, CS,$ and CR

When the driver intends to maintain a constant speed (i.e., $s \in \{CL, CS, CR\}$), there is no concept of acceleration maximisation or deceleration minimisation. Therefore, we specify a simple, constants-only discrete choice model for the steering angle choice for each of the intents $CL, CS,$ or CR. Further, since there is no intent to accelerate or decelerate, the extent of intended acceleration is zero (that is, the modelled acceleration, $A^s_{q(t+\Delta t)} = 0 \;\forall s \in \{CL, CS, CR\}$), albeit the observed acceleration values may not be zero as the drivers unintendedly accelerate or decelerate by a small amount.

6.2.2.5 Difference between observed and planned acceleration values: Execution error

The acceleration values obtained from the models for $A^s_{q(t+\Delta t)}$ in the above sections (6.2.2.2 through 6.2.2.4) may be viewed as planned acceleration values. These values are quite likely to be different from the executed (or, observed) acceleration values in the field (trajectory data).

This difference is viewed as the driver's execution error attributable to the vehicle machine capabilities and the driver's driving experience. Specifically, the execution error for any of the intents is expressed as:

$$\eta^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - A^{s}_{q(t+\Delta t)} \; \forall s \in J \tag{6.7}$$

where, $\eta^{s}_{q(t+\Delta t)}$ is a random term representing the execution error for intent s, $A^{s}_{q(t+\Delta t)}$ is the modelled acceleration for that intent, and $A_{q(t+\Delta t)}$ is the observed acceleration.

Note that the 2D movement of the SV can potentially be constrained due to physical boundaries such as road edges. As a result, when the SV is at the laterally left or right extreme ends of the road, we constrain the choice set ($G^{s}_{q(t+\Delta t)}$) for all the steering angle choice models. For example, it is presumed that the subject vehicle does not have an option of steering along l_3 (r_3) when it is at the extreme left (right) end of the road. The locations of extreme ends are determined based on empirical data. It is observed in the data that the SVs did not choose angle l_3 when the lateral distance between the left edge of the SV and the left edge of the road was less than 0.80 m. Similarly, the angle r_3 was not chosen by the SVs when the lateral distance between the right edge of the SV and the right edge of the road was less than 1.64 m. Hence, we considered the SV as being at the extreme left (right) end when the lateral distance between the left (right) edge of the road and the left (right) edge of the SV was less than 0.80 m (1.64 m).

6.2.3 Likelihood Function Formulation

To derive the likelihood function, we mapped drivers' manoeuvres observed from trajectory data (i.e., the executed acceleration and steering angle) to their intents, as shown in Figure 6.4.

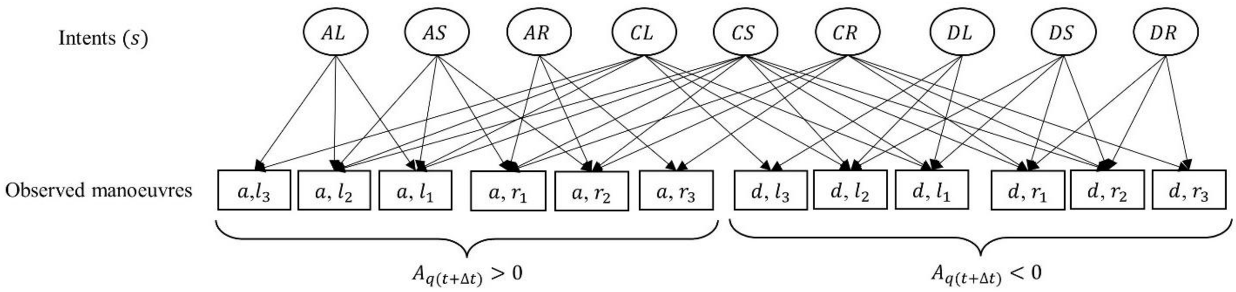


Figure 6.4 Mapping of drivers' observed actions to the latent intents

Following this mapping, the likelihood function for the observed actions of the driver of vehicle q at time $(t+\Delta t)$ is written as:

$$L_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\right)=\left[\sum_{s\in\{AL,AS,CL,CS\}}P_{q(t+\Delta t)}(s)P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{\gamma_{q(t+\Delta t)}}\left[\sum_{s\in\{DL,DS,CL,CS\}}P_{q(t+\Delta t)}(s)P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{1-\gamma_{q(t+\Delta t)}}\quad\forall i\in\{l_1,l_2\}$$

$$L_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\right)=\left[\sum_{s\in\{AR,AS,CR,CS\}}P_{q(t+\Delta t)}(s)P_{qt}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{\gamma_{q(t+\Delta t)}}\left[\sum_{s\in\{DR,DS,CR,CS\}}P_{q(t+\Delta t)}(s)P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{1-\gamma_{q(t+\Delta t)}}\quad\forall i\in\{r_1,r_2\}$$

$$L_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\right)=\left[\sum_{s\in\{AL,CL\}}P_{q(t+\Delta t)}(s)P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{\gamma_{q(t+\Delta t)}}\left[\sum_{s\in\{DL,CL\}}P_{q(t+\Delta t)}(s)P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{1-\gamma_{q(t+\Delta t)}}\quad\forall i\in\{l_3\}$$

$$L_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\right)=\left[\sum_{s\in\{AR,CR\}}P_{q(t+\Delta t)}(s)P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{\gamma_{q(t+\Delta t)}}\left[\sum_{s\in\{DR,CR\}}P_{q(t+\Delta t)}(s)P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},\theta_i\middle|s\right)\right]^{1-\gamma_{q(t+\Delta t)}}\quad\forall i\in\{r_3\}$$

(6.8)

In this expression, $\gamma_{q(t+\Delta t)}=1$ if $A_{q(t+\Delta t)}>0$, otherwise $\gamma_{q(t+\Delta t)}=0$. $P_{q(t+\Delta t)}(s)$ is the logit expression in Eq. (6.1).

6.2.3.1 Conditional likelihood expressions $P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\middle|s\right)$ for different intents

$P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\middle|s\right)\forall s\in\{AL,AR,DL,DR\}$: From Eq. (2),

$$\begin{aligned}P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\middle|s\right)&=\mathrm{P}\left(A^s_{q(t+\Delta t)i}\geq A^s_{q(t+\Delta t)j}\ \forall j\in G^s_{q(t+\Delta t)},\ \eta^s_{q(t+\Delta t)}=A_{q(t+\Delta t)}-A^s_{qi(t+\Delta t)}\right)\\&=\mathrm{P}\left(a^s_{q(t+\Delta t)i}+\varepsilon^s_{qti}\geq a^s_{q(t+\Delta t)j}+\varepsilon^s_{qtj}\ \forall j\in G^s_{q(t+\Delta t)},\ \eta^s_{q(t+\Delta t)}=A_{q(t+\Delta t)}-a^s_{q(t+\Delta t)i}-\varepsilon^s_{qti}\right)\\&=\mathrm{P}\left(\varepsilon^s_{qtj}-\varepsilon^s_{qti}\leq a^s_{q(t+\Delta t)i}-a^s_{q(t+\Delta t)j}\ \forall j\in G^s_{q(t+\Delta t)},\ \eta^s_{q(t+\Delta t)}+\varepsilon^s_{qti}=A_{q(t+\Delta t)}-a^s_{q(t+\Delta t)i}\right)\end{aligned}$$

(6.9)

By re-expressing some of the above terms as follows: $\varepsilon^s_{qtj}-\varepsilon^s_{qti}=\tilde{\varepsilon}^s_{qtji}$, $a^s_{q(t+\Delta t)i}-a^s_{q(t+\Delta t)j}=\tilde{a}^s_{q(t+\Delta t)ij}$, and $\eta^s_{q(t+\Delta t)}+\varepsilon^s_{qti}=\tilde{\eta}^s_{q(t+\Delta t)}$, Eq. (6.9) can be simplified as below:

$$P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)},i\middle|s\right)=\mathrm{P}\left(\tilde{\eta}^s_{q(t+\Delta t)}=A_{q(t+\Delta t)}-a^s_{q(t+\Delta t)i}\right)\times\mathrm{P}\left(\tilde{\varepsilon}^s_{qtji}<\tilde{a}^s_{q(t+\Delta t)ij}\ \forall j\in G^s_{q(t+\Delta t)}\middle|\tilde{\eta}^s_{q(t+\Delta t)}=A_{q(t+\Delta t)}-a^s_{q(t+\Delta t)i}\right)$$

(6.10)

We assume that ε^s_{qti} are IID normally distributed error terms with mean zero and variance one and $\eta^s_{q(t+\Delta t)}$ are assumed as IID normally distributed error terms with mean zero and variance $(\sigma^s_\eta)^2$. Then, $\tilde{\eta}^s_{q(t+\Delta t)}$ would be normally distributed with mean zero and variance $\left(\sigma^s_\eta\right)^2+1$. As a result, $\mathrm{P}\left(\tilde{\eta}^s_{q(t+\Delta t)}=A_{q(t+\Delta t)}-a^s_{q(t+\Delta t)i}\right)$ becomes a normal probability density function (pdf) as below:

$$\mathrm{P}\left(\tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i}\right) = \frac{1}{\sqrt{\left(\sigma^{s}_{\eta}\right)^{2}+1}}\,\phi\left(\frac{A_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i}}{\sqrt{\left(\sigma^{s}_{\eta}\right)^{2}+1}}\right) \tag{6.11}$$

To simplify the conditional likelihood expression in Eq. (6.10), we utilise a property of multivariate normal (MVN) distribution (Tong, 1990) that the distribution of $\tilde{\varepsilon}^{s}_{qtji}\ \forall j \in G^{s}_{q(t+\Delta t)}, j \neq i$ given $\tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i}$, is another MVN distribution as given below:

$$(\mathbf{A} \mid \mathbf{B} = \mathbf{b}) \sim N(\bar{\boldsymbol{\mu}}, \boldsymbol{\Omega}), \text{ where } \bar{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}^{-1}_{22}(\mathbf{b} - \boldsymbol{\mu}_2), \text{ and } \bar{\boldsymbol{\Omega}} = \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}^{-1}_{22}\boldsymbol{\Omega}_{21} \tag{6.12}$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{b}$, and $\mathbf{a}$ are vectors of $\tilde{\varepsilon}^{s}_{qtji}$, $\tilde{\eta}^{s}_{q(t+\Delta t)}$, $(a_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i})$, and $\tilde{a}^{s}_{q(t+\Delta t)ij}$, respectively. Now, the conditional likelihood expression in Eq. (6.10), can be written as:

$$\begin{aligned} \mathrm{P}\left(\tilde{\varepsilon}^{s}_{qtji} < \tilde{a}^{s}_{q(t+\Delta t)ji} \middle| \tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i}\right) &= \mathrm{P}(\mathbf{A} < \mathbf{a} \mid \mathbf{B} = \mathbf{b}) \\ &= \mathrm{P}(\mathbf{C} < \mathbf{a}), \end{aligned} \tag{6.13}$$

where $\mathbf{C}$ is an MVN random variable with a mean vector $\bar{\boldsymbol{\mu}}$ and variance-covariance matrix $\bar{\boldsymbol{\Omega}}$ as in Eq. (6.12). $\Pr(\mathbf{C} < \mathbf{a})$ does not have a closed-form expression. The current study employs a matrix-based method for the analytic approximation of the MVN cumulative distribution function provided by Bhat (2018).

$\underline{P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)}, i \middle| s\right) \forall s \in \{AS, DS\}}$: Following Section 6.2.2.3 for intents $s \in \{AS, DS\}$,

$$\begin{aligned} P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)}, i \middle| s\right) &= P_{q(t+\Delta t)}(i \mid s) \times \mathrm{P}\left(\eta^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - A^{s}_{qi(t+\Delta t)}\right) \forall s \in \{AS, DS\} \\ &= P_{q(t+\Delta t)}(i \mid s) \times \mathrm{P}\left(\eta^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i} - \varepsilon^{s}_{qti}\right) \forall s \in \{AS, DS\} \\ &= P_{q(t+\Delta t)}(i \mid s) \times \mathrm{P}\left(\tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i}\right) \forall s \in \{AS, DS\} \end{aligned} \tag{6.14}$$

As the choice of angle i conditional on intents $s \in \{AS, DS\}$ is modelled using a simple discrete choice framework with only constants in the specification, $P_{qt}(i \mid s)$ in the above equation takes a multinomial probit type discrete choice model form. $\Pr\left(\tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - a^{s}_{q(t+\Delta t)i}\right)$ is a normal pdf expressed as in Eq. (6.11).

$\underline{P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)}, i \middle| s\right) \forall s \in \{CL, CS, CR\}}$: Following Section 2.2.3 for intents $s \in \{CL, CS, CR\}$,

$$\begin{aligned} P_{q(t+\Delta t)}\left(A_{q(t+\Delta t)}, i \middle| s\right) &= P_{q(t+\Delta t)}(i \mid s) \times \mathrm{P}\left(\eta^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)} - \varepsilon^{s}_{qti}\right) \forall s \in \{CL, CS, CR\} \\ &= P_{q(t+\Delta t)}(i \mid s) \times \mathrm{P}\left(\eta^{s}_{q(t+\Delta t)} + \varepsilon^{s}_{qti} = A_{q(t+\Delta t)}\right) \forall s \in \{CL, CS, CR\} \\ &= P_{q(t+\Delta t)}(i \mid s) \times \mathrm{P}\left(\tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)}\right) \forall s \in \{CL, CS, CR\} \end{aligned} \quad (6.15)$$

Since a simple, constants-only discrete choice model is used for the steering angle choice for each of the intents $CL, CS,$ or CR, $P_{q(t+\Delta t)}(i \mid s)$ in the above equation takes the multinomial probit type discrete choice model form. Note that we estimate the mean (μ^{C}) and standard deviation σ^{C}_{η} of error terms $\eta^{s}_{q(t+\Delta t)}$ for all $s \in \{CL, CS, CR\}$. $\mathrm{P}\left(\tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)}\right)$ is a normal pdf expressed as below:

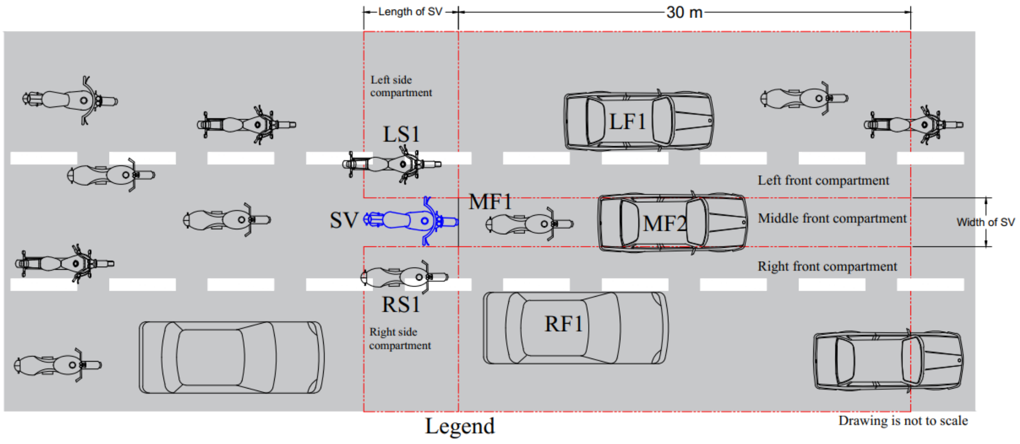
$$\mathrm{P}\left(\tilde{\eta}^{s}_{q(t+\Delta t)} = A_{q(t+\Delta t)}\right) = \frac{1}{\sqrt{\left(\sigma^{C}_{\eta}\right)^{2} + 1}} \phi\left(\frac{A_{q(t+\Delta t)} - \mu^{C}}{\sqrt{\left(\sigma^{C}_{\eta}\right)^{2} + 1}}\right) \quad (6.16)$$

6.3 EMPIRICAL MODEL

The above described framework is applied to develop an empirical model of motorised two-wheeler driver behaviour using a vehicle trajectory dataset from an urban arterial road in Chennai, India (2015). In this section, we first describe the traffic environment variables used to explain the driver behaviour and then present the empirical findings.

6.3.1 Traffic Environmental Variables in the Latent Class Model for Higher-Level Intents

It is assumed that the subject vehicle (SV)'s higher-level decisions (intents) are influenced by multiple vehicles around the subject vehicle. Following the analysis done in Chapter 3, the MVA effect is incorporated using the concept of an *influence zone* (see Figure 6.5), where vehicles within the influence zone around an SV are assumed to influence its behaviour. In Figure 6.5, the subject vehicle (SV) is in blue colour.



SV : Subject vehicle	RF1 : First vehicle in right front compartment
MF1: First vehicle in middle front compartment	LS1 :First vehicle in left side compartment
MF2 : Second vehicle in middle front compartment	RS1 :First vehicle in right side compartment
LF1 : First vehicle in left front compartment	:Influence zone

Figure 6.5 Structure of a rectangular influence zone around a subject vehicle

As described in Chapter 3, the influence zone around the SV is divided into five compartments. The labelling of the compartments is according to the position of the compartment with respect to the SV. Space directly ahead of the subject vehicle is called the middle front (MF) compartment, the space ahead to the left of the subject vehicle is labelled the left front (LF) compartment (similarly, the space ahead to the right of the subject vehicle is called the RF compartment), and the adjacent space to the left of the subject vehicle is called the left side (LS) compartment (similarly adjacent space to the right of the subject vehicle is called the RS compartment). The vehicles in these compartments are labelled as in the legend of Figure 6.5. Specifically, vehicles in the influence zone other than the SV are assigned a number based on their proximity to the SV and their compartment. For example, the closest vehicle in the MF compartment is labelled MF1. Next, traffic environment variables (such as space gaps and relative speeds) are calculated with respect to each of these surrounding vehicles for use in the latent class model.

6.3.2 Calculation of Attributes Used in the Lower-Level Model for the Extent of Acceleration

6.3.2.1 Spacing (ΔX_{qti}) calculation

The distance between the subject vehicle (SV) and the lead vehicle (LV) along an angular direction i, measured along that direction, is considered as the space gap ΔX_{qti} for that direction. To understand this, consider that the subject vehicle (SV) has three LVs, as in Figure

6.6. The points on LV that are at minimum distance from the SV are identified for each angle (only if the LV is present in that angle). Then the distances between these points and the subject vehicle are calculated. If no LV is present along a given angular direction i, then the distance up to the curb is considered as spacing for that direction and represented as ΔX_{qti}^{curb} in Eq. (6.3) . Note that, for ease of presentation, the subscript qt is suppressed when spacing is denoted in Figure 6.6.

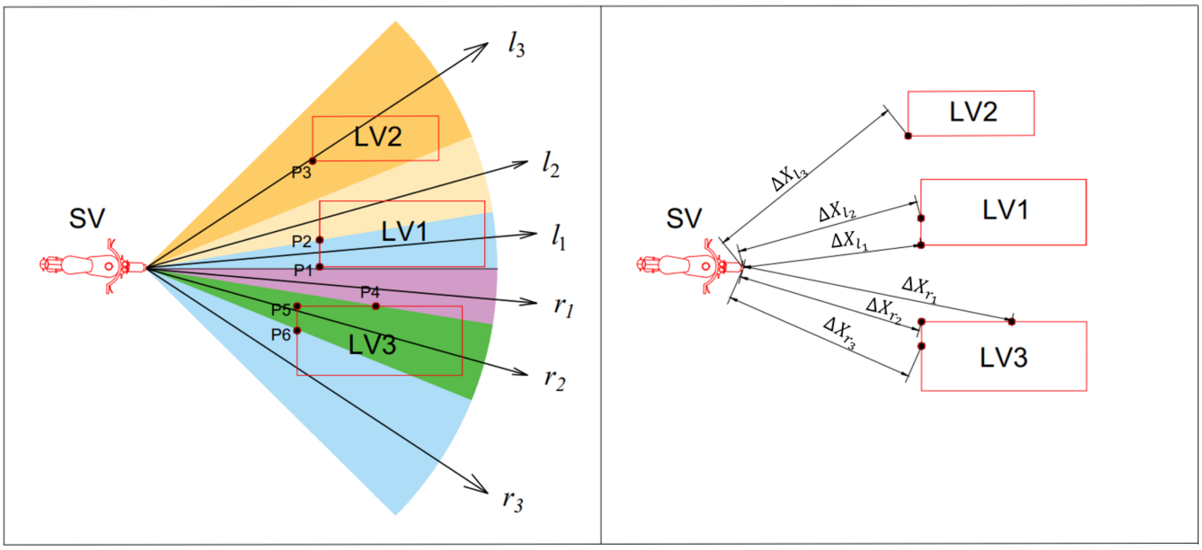


Figure 6.6 Spacing (ΔX_{qti}) between subject vehicle (SV) and lead vehicle (LV) along direction i

6.3.2.2 Relative velocity (ΔV_{qti}) calculation

Relative velocity between the SV and an LV along the direction i is defined as the difference between the LV velocity and the SV velocity, after projecting both velocities along that direction. Figure 6.7 depicts the calculation of relative velocity. In this figure, V_{SV} is the velocity vector of the SV calculated using the observed longitudinal speed $(V_{SV,X})$ and lateral speed $(V_{SV,Y})$. The angle of V_{SV} with respect to the longitudinal axis is $\delta_{SV} = \tan^{-1}\left(\frac{V_{SV,Y}}{V_{SV,X}}\right)$.

The projection of V_{SV} along the direction i is $V_{SV}\cos\left(|\delta_{SV} - \theta_i|\right)$. Similarly, the projection of the velocity of LV along i is $V_{LV}\cos\left(|\delta_{LV} - \theta_i|\right)$. The relative velocity between the two vehicles along the direction i is the difference between the projections of LV velocity and the SV velocity along i, as below:

$$\Delta V_i = V_{LV}\cos\left(|\,\delta_{LV} - \theta_i\,|\right) - V_{SV}\cos\left(|\,\delta_{SV} - \theta_i\,|\right) \tag{6.17}$$

Note that the subscript qt is suppressed in the above equation and in Figure 6.7 for ease of presentation. In Figure 6.7, it is assumed that the LV is present ahead of the SV along the direction i. Hence, ΔV_i is only calculated when the LV is present along i. The same procedure is followed for all $i \in I = \{l_1, l_2, l_3, r_1, r_2, r_3\}$.

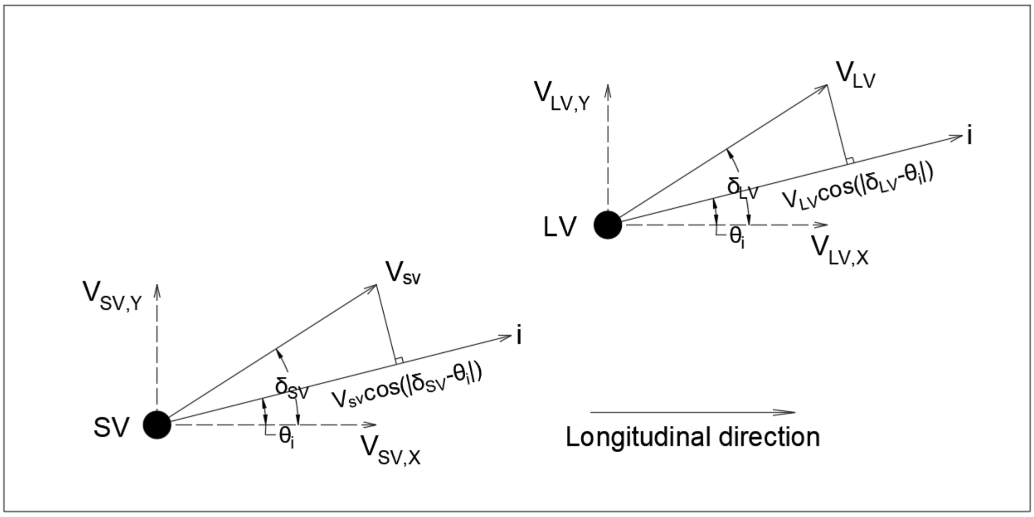


Figure 6.7 Speeds of SV and LV along the direction i

6.4 EMPIRICAL MODEL RESULTS AND DISCUSSION

6.4.1 Estimation Results of Latent Class Model for Higher-level Intents

Table 6.1 shows the estimation results of the latent class model for drivers' higher-level intents. Recall that we had identified nine possible intents (S). To avoid proliferation of parameters, the effect of traffic environment variables was not specified separately in each intent. Instead, a dimension-wise specification was employed, where the effect of each variable was introduced separately for the acceleration intents and the steering direction intents. To specify the variable effects on acceleration intents, maintaining a constant speed was considered the base category. For the variable effects on steering direction intents, steering in the straight (longitudinal) direction was considered the base category. Further, to better interpret some of the empirical model results, it is worth noting here that vehicular traffic in India follows a "keep left" driving policy, where the left side of the travel direction is the curb side, and the right side of the travel direction is the median side (if the road is median separated).

The parameter estimates of SV's longitudinal speed variable suggest, as expected, that the drivers of vehicles travelling at slower speeds are more (less) likely to intend to accelerate (decelerate) than drivers travelling at faster speeds. In the context of steering direction, motorised two-wheeler drivers in HD traffic conditions, when travelling at higher longitudinal speeds, are more (less) likely to intend to steer left (right) of the longitudinal direction. This finding can be explained based on two aspects of two-wheeler driver behaviour typical to Indian traffic conditions. First, two-wheelers on Indian roads tend to travel on the left side of the road, segregated from larger vehicles such as cars that travel on the right side (Transport Department, Government of India, 1989). Second, since a two-wheeler typically travels on the left side of the road, its manoeuvre to overtake a slow-moving vehicle ahead involves the following steps: (1) steer to the right of the longitudinal direction (after finding a gap to do so), (2) speed up (while steering to the right and while steering back to the longitudinal direction) to get past the slow-moving vehicle, and (3) steer to the left to get back to the left side of the road (Transport Department, Government of India, 1989). Therefore, two-wheelers at high speeds are more likely (than slower two-wheelers) to be those that are steering to the left of the road to complete an overtaking manoeuvre. Besides, two-wheelers are more likely to mitigate the heightened risk of higher speeds by moving to the left side of the road where there are more vehicles of the same type. Further, vehicles at higher speeds have less incentive (when compared to those at slower speeds) to steer right to initiate an overtaking manoeuvre.

Next, we turn to the influence of traffic environment variables in the MF compartment. The space gap with respect to the MF1 vehicle shows a strong negative influence on the intent to decelerate. However, there is no significant differential effect of this space gap between the intent to accelerate or to remain in same speed. Furthermore, space gap with respect to MF1 does not show a significant influence on the steering direction intents. On the other hand, the relative speed with respect to MF1 appears to be a stronger stimulus (than the space gap), with a significant influence on the intent to accelerate (+ve influence), the intent to decelerate (-ve influence), and the intent to steer right of the longitudinal direction (-ve influence). The influences of relative speed on SV's likelihood of accelerating or decelerating are well documented in the literature (Koutsopoulos and Farah, 2012), albeit the influence on the steering direction has not been well explored earlier. Intuitively, an SV with an LV longitudinally ahead of it at a higher speed (than the SV) has little motivation to steer in the right direction. In contrast, if the MF1 is driving slower than the subject vehicle, the SV's driver will want to overtake the lead vehicle; hence the driver will be more inclined to steer right rather than move in the longitudinal direction. This result corroborates earlier findings in the literature (Chandra and Shukla, 2012; Asaithambi and Shravani, 2017) that in HD traffic conditions of Indian cities, when drivers have the option to go past the slow-moving lead vehicle, they tend to overtake it from the right side.

In addition to the first LV (MF1) in the MF compartment, the empirical results suggest an influence of the second LV (MF2) as well. Specifically, the presence of MF2 (along with MF1) might make the SV driver feel constrained and motivate them to overtake by steering to the right of the longitudinal direction. At the same time, presence of a large space gap between SV and MF2 would reduce such an urge. The influence of relative speed with respect to MF2 on the SV driver's intent to accelerate or decelerate is similar to that of MF1, though the influence of MF2 is weaker than that of MF1. While earlier studies (e.g., Hoogendoorn and Ossen, 2006; Hoogendoorn et al., 2006; and Zhang, 2014) recognize that the relative influences of MF2 and MF1 on the longitudinal acceleration behaviour of vehicles, we are not aware of studies that highlight the influence of MF2 on a subject vehicle's 2D movement.

The above-discussed influences of MF2 offer evidence of the MVA effect on drivers' higher-level intents. Along the same lines, vehicles in the LF, RF, LS, and RS compartments have an influence too. In this context, the absence of other vehicles in the RF and RS compartments shows a greater likelihood of drivers' intent to steer to the right of the longitudinal direction – perhaps because of the opportunity provided by the empty space on the right side of the SV to possibly overtake an LV in the MF compartment (Chandra and Shukla, 2012; Asaithambi and Shravani, 2017). Similarly, the absence of vehicles in the LF and LS compartments seem to increase drivers' intent to steer to the left. By the same token, the presence of a vehicle in these compartments would reduce the intent to steer left. However, if a vehicle is present in the LF compartment, a high space gap between it and the SV would again increase the intent to steer left, *ceteris paribus*. Further, if the LF1 moves faster than the SV, it seems to reduce the intent of the SV's driver to decelerate or to steer in the right direction. Interestingly, however, the space gap with respect to the LV labelled RF1 does not have an influence on any of the intents. At the same time, the relative speed between the SV and RF1 influences the driver's intents along both the acceleration and steering direction dimensions. Finally, the lateral space available to the left of the SV – either in the form of space between SV and LS1 or between SV and the left edge of the road – influences the driver's intents. Specifically, SV drivers show higher intents to accelerate and to steer left when they find large lateral space gaps to their left side. This is consistent with our expectations since, in HD traffic scenarios (especially in the Indian driving context) two-wheelers prefer to be on the left side of the road. Overall, the latent class model suggests behaviourally plausible influences of various surrounding vehicles on the SV driver's intents, besides highlighting the MVA effect. In addition, the empirical results highlight driving behaviours typical to motorized two-wheelers in HD traffic streams in Indian cities.

6.4.2 Models for Lower-Level Decisions (Steering Angle Choice and Acceleration) Conditional on Higher-Level Intents

Table 6.2 reports the estimation results of the models for lower-level decisions – choice of the specific steering angle and the extent of acceleration or deceleration – conditional on the higher-level intents. Several interesting behavioural patterns surface from these models and their comparison with the findings of the earlier-discussed model for higher-level intents. First, the gap between the desired speed and the SV speed appears to be the primary variable influencing the lower-level decisions once the higher-level intents are made (that too only when the intents are AL, DL, and DR). Second, the only other traffic environment variable that shows an influence on the lower-level decisions is the relative velocity between the SV and the lead vehicle – that too, only when the intent is AR (i.e., accelerate and steer right of longitudinal direction). Third, space gaps do not show significant influence on lower-level decisions once higher-level intents are made (in which space gaps did show a strong influence). In fact, we see that no traffic environment variable has a significant influence on the extent of acceleration (deceleration) when the drivers' intent is AS (DS). All these results suggest that the influence of traffic environment variables on lower-level, tactical decisions is not as strong as that found on higher-level decisions (intents). A plausible explanation for such results is that drivers invest greater cognitive resources in making their higher-level, strategic intents than what they invest in making the lower-level, tactical decisions. Once they make their higher-level intents based on a very careful evaluation of the traffic environment, it may be that they make the lower-level decisions based on their driving experience as opposed to an equally careful evaluation of the traffic environment.

The influence of the relative velocity variable conditional on the AR intent is perhaps because the SV is attempting to overtake the vehicle ahead. In Indian traffic conditions, drivers overtake from the right side of vehicles ahead of them, and existing literature suggests that relative velocity is an essential factor in overtaking manoeuvres (Chandra and Shukla, 2012; Asaithambi and Shravani, 2017).

Note that we estimated constants (α_{s1}) for all steering angle choice alternatives in the models specific to intents AL, AR, DL, and DR. In fact, we constrained the constants to be the same for all steering angle alternatives for a given intent. This is unlike the usual discrete choice models where the constant for at least one alternative should be normalised for identification purposes. We estimated the same constant in all alternatives because a unified model of Eq. (2), combined with Eq. (3), is used for both the choice of the steering angle (from a set of discrete alternatives) and the extent of acceleration or deceleration (a continuous value). In this model, the observed steering angle choice informs the estimation of coefficients on the traffic environment variables along each angular direction, while the observed acceleration value informs the estimation of the constant. Such a constant helps in fixing the location (on the real line) of the extent of acceleration obtained from Eq. (6.3) for intents AL, AR, DL, and DR.

In the lower-level decision models for intents AS and DS, the reader will note two sets of constants – one in the steering angle choice model and another in the model for the extent of acceleration. This is because for these intents, as discussed in Section 6.2.2.3, the choice of the steering angle is not coupled with the extent of acceleration. The steering angle is modelled using a simple, constants only discrete choice model. The positive value of the constants estimated specific to angular directions l_1 and r_1 in the model for DS intent, with the base categories l_2 and r_2, suggest that drivers are more likely to steer along the l_1 or r_1 directions (than the l_2 or r_2 directions), when their intent was to keep straight. Next, we turn to the constants in Eq. (3) used for the extent of acceleration in the models for AS and DS intents. These constants indicate that the expected acceleration (deceleration) value when the intent is AS (DS) is 1.039 m/s^2 (-1.123 m/s^2).

In the lower-level decision models for intents CL, CS, and CR, as discussed in Section 6.2.2.4, the expected value of acceleration is zero (because the driver intended to be in a constant speed). Therefore, a simple, constants only discrete choice model is used for the steering angle choice.

Table 6.1 Estimation results of the latent class model for higher-level decisions (intents)

Traffic environment variable	Acceleration intents		Steering direction intents	
	Accelerate (A)	Decelerate (D)	Steer left (L)	Steer right (R)
Constant	2.976 (9.05)	-5.816 (-13.98)	-4.222 (-8.48)	2.147 (5.39)
Speed of the subject vehicle (m/s)	-0.520 (-15.83)	0.427 (12.45)	0.056 (2.05)	-0.344 (-9.45)
Traffic environment variables with respect to vehicles in MF compartment				
Space gap in longitudinal direction with respect to MF1 (m)	--	-0.037 (-5.20)	--	--
Relative speed in longitudinal direction with respect to MF1 (m/s)	0.166 (8.32)	-0.080 (-3.67)	--	-0.142 (-5.72)
Subject vehicle has two or more lead vehicles in MF compartment	--	--	--	0.789 (1.95)
Space gap in longitudinal direction with respect to MF2 (m)	--	--	--	-0.032 (-1.74)
Relative speed in longitudinal direction with respect to MF2 (m/s)	0.122 (3.48)	-0.037 (-1.29)	--	--
Subject vehicle does not have lead vehicles in RF and RS compartments	--	--	--	0.179 (1.38)
Subject vehicle does not have lead vehicles in LF and LS compartments	--	--	0.912 (4.54)	--
Traffic environment variables with respect to vehicles in LF and RF compartments				
Space gap in longitudinal direction with respect to LF1 (m)	--	--	0.044 (6.73)	--
Relative speed in longitudinal direction with respect to LF1 (m/s)	--	-0.066 (-3.23)	--	-0.056 (-2.40)
Space gap in longitudinal direction with respect to RF1 (m)	--	--	--	--
Relative speed in longitudinal direction with respect to RF1 (m/s)	0.057 (3.22)	-0.060 (-3.00)	-0.046 (-1.91)	0.024 (1.05)
Traffic environment variables with respect to vehicles in LS and RS compartments				
Space gap in lateral direction with respect to LS1 (m)	0.182 (4.99)	--	--	--
Relative speed in lateral direction with respect to LS1 (m/s)	--	--	--	--
Space gap in lateral direction with respect to RS1 (m)	--	--	--	--
Relative speed in lateral direction with respect to RS1 (m/s)	--	--	--	--
Space gap between left edge of the SV and left edge of the road (m)	0.132 (5.19)	-0.281 (-10.39)	0.118 (3.24)	--

-- the corresponding parameter was dropped from the specification as it was found to be statistically insignificant

Table 6.2 Estimation results of the models for lower-level decisions

Variable	**Estimates (t-stats)**	**Variable**	**Estimates (t-stats)**
When AL is higher-level intent (steering angle alternatives: $i \in \{l_1, l_2, l_3\}$)		**When DL is higher-level intent (steering angle alternatives:** $i \in \{l_1, l_2, l_3\}$)	
Steering angle choice and acceleration model		Steering angle choice and acceleration model	
Constants (in all alternatives)	0.248 (0.68)	Constant (in all alternatives)	-2.769 (-29.48)
Difference between desired speed and speed of the subject vehicle (m/s)	0.085 (1.95)	Difference between desired speed and speed of the subject vehicle (m/s)	0.113 (4.32)
Difference between available space gap and desired space gap (m)	--	Difference between available space gap and desired space gap (m)	--
Relative velocity with respect to lead vehicle (m/s)	--	Relative velocity with respect lead vehicle (m/s)	--
When AS is higher-level intent (steering angle options: $i \in \{l_1, l_2, r_1, r_2\}$)		**When DS is higher-level intent (steering angle options:** $i \in \{l_1, l_2, r_1, r_2\}$)	
Steering angle choice model		Steering angle choice model	
Constants specific to l_1 and r_1 (l_2 and r_2 are base alternatives)	-0.037 (-0.57)	Constants specific to l_1 and r_1 (l_2 and r_2 are base alternatives)	0.217 (3.04)
Acceleration model		Acceleration model	
Constant	1.039 (29.69)	Constant	-1.123 (-27.73)
Difference between desired speed and speed of the subject vehicle (m/s)	--	Difference between desired speed and speed of the subject vehicle (m/s)	--
Difference between available space gap and desired space gap (m)	--	Difference between available space gap and desired space gap (m)	--
Relative velocity with respect to lead vehicle (m/s)	--	Relative velocity with respect lead vehicle (m/s)	--
When AR is higher-level intent (steering angle alternatives: $i \in \{r_1, r_2, r_3\}$)		**When DR is higher-level intent (steering angle alternatives:** $i \in \{r_1, r_2, r_3\}$)	
Steering angle choice and acceleration model		Steering angle choice and acceleration model	
Constants (in all alternatives)	0.599 (4.44)	Constants (in all alternatives)	-2.885 (-18.38)
Difference between desired speed and speed of the subject vehicle (m/s)	--	Difference between desired speed and speed of the subject vehicle (m/s)	0.147 (3.78)
Difference between available space gap and desired space gap (m)	--	Difference between available space gap and desired space gap (m)	--
Relative velocity with respect to lead vehicle (m/s)	0.079 (4.08)	Relative velocity with respect lead vehicle (m/s)	--
When CL is higher-level intent (steering angle alternatives: $i \in \{l_1, l_2, l_3\}$ **)**			
Steering angle choice model			
Constants specific to l_1 and l_2 (l_3 is base alternative)	0.650 (6.45)		
When CS is higher-level intent (steering angle options: $i \in \{l_1, l_2, r_1, r_2\}$)			
Steering angle choice model			
Constants specific to l_1 and r_1 (l_2 and r_2 are base alternatives)	0.493 (14.73)		
When CR is higher-level intent (steering angle alternatives: $i \in \{r_1, r_2, r_3\}$ **)**			
Steering angle choice model			
Constants specific to r_1 and r_2 (r_3 is base alternative)	0.589 (5.59)		
Scale parameters of execution error terms $\eta^{s}_{q(t+\Delta t)}$		**Mean parameters of execution error terms** $\eta^{s}_{q(t+\Delta t)}$	
σ_η^{AL}	0.385 (1.81)	μ^A when the intent is to accelerate (AL, AS, AR)	0.00 (Fixed)
σ_η^{AS}	0.447 (17.43)	μ^D when the intent is to decelerate (DL, DS, DR)	0.00 (Fixed)
σ_η^{AR}	0.274 (2.15)	μ^C when the intent is to be in constant speed (CL, CS, CR)	-0.062 (-6.00)
$\sigma_\eta^{CL} = \sigma_\eta^{CS} = \sigma_\eta^{CR} = \sigma_\eta^{C}$	0.453 (58.04)		
σ_η^{DL}	0.345 (2.09)		
σ_η^{DS}	0.472 (16.87)		
σ_η^{DR}	0.372 (1.29)		
Number of parameters		52	
Log-likelihood at convergence		-20213.57	

-- the corresponding parameter was dropped from the specification as it was found to be statistically insignificant

6.4.3 Parameters of the Execution Error Term

As discussed in Section 6.2.2.5, the proposed framework recognises execution error ($\eta^{s}_{q(t+\Delta t)}$) as the difference between executed acceleration and modelled acceleration. In all lower-level decision models conditional on the higher-level intents being acceleration or deceleration, the scale parameters are estimated while fixing the mean parameters (see the last set of rows in Table 6.2). If the drivers intended to remain at a constant speed, both the scale and mean parameters are estimated for the execution error term. The mean parameter indicates a systematic bias away from zero acceleration when the drivers intended to remain at constant speed. The estimate of this mean parameter takes a value of -0.06, which implies that even if the drivers intended to remain at a constant speed, their execution is biased toward an average deceleration of -0.06 m/s^2. This may be because drivers are conservative and prefer to err on the side of safety.

6.4.4 Model Comparison

To evaluate the benefits of assuming a two-step decision process, where drivers' higher-level intents precede their tactical decisions of the specific steering angle and the extent of acceleration or deceleration, we estimated three reduced-form models that do not have a latent class layer for the intents. The first of these reduced-form models (labelled RM1), as shown in Figure 6.8(a), treats the observed driving decisions directly as a result of steering angle choice and the extent of acceleration or deceleration. This model involves a multinomial probit-based discrete choice framework for the steering angle choice (from the directions $l_1, l_2, l_3, r_1, r_2, r_3$) and a regression model for the extent of acceleration or deceleration considering the effect of the immediate lead vehicle. The second reduced-form model (RM2) formulates a two-stage procedure, albeit without a latent class framework, as shown in Figure 6.8(b). Specifically, the decisions to accelerate or decelerate (there is no decision to remain at constant speed) and the decisions to steer left or right (without an option to keep straight) are represented as first-stage decisions. In the second stage, the choice of the specific angle and the extent of acceleration are modelled separately conditional on the first-stage decisions. The third reduced-form model (RM3) formulates another multi-stage decision process, as shown in Figure 6.8(c) (again without the latent class framework). Similar to RM2, in this formulation, the decisions to accelerate or decelerate and the decisions to steer left or right are viewed as the first stage decisions. Conditional on these decisions, the choice of the specific angle and the extent of acceleration are modelled separately.

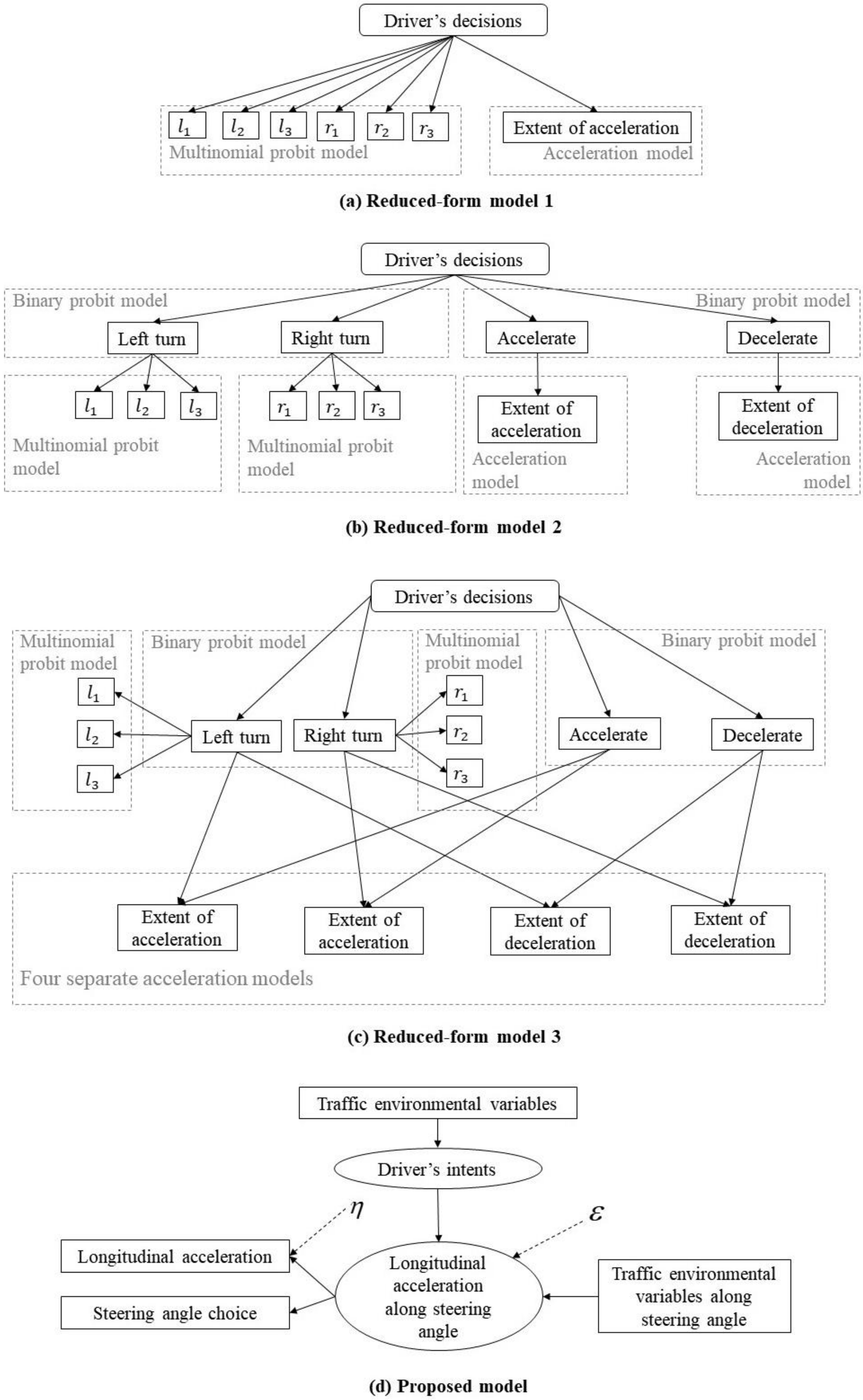


Figure 6.8 Structure of the reduced-form models and the proposed model

The main difference between the reduced-form models and the proposed model is that the latter uses a latent class framework to consider the possibility that the drivers may have intended to be at a constant speed and/or to keep straight. Besides, the choice of the specific steering angle is based on a defensible theory that drivers choose an angle that allows them to proceed at maximum (minimum) possible acceleration (deceleration) if they had intended to accelerate (decelerate).

Table 6.3 presents the goodness-of-fit metrics of all the estimated models to the estimation data. Since the reduced-form models are not special cases of the proposed model, the typically used likelihood ratio test cannot be used for comparing the models. Hence, Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) metrics are employed to compare the data fit of different models. As can be observed from the AIC and BIC values in the table, the proposed model outperforms all the three reduced form models in terms of model fit. These results suggest the importance of the latent class framework for considering the possibility of intents unobservable from trajectory datasets.

Table 6.3 Model comparison

Goodness of fit measures	Proposed model	Reduced-form model (RM) 1	Reduced-form model (RM) 2	Reduced-form model (RM) 3
No. of cases	7748	7748	7748	7748
No. of parameters	52	16	41	52
Log-likelihood	-20213.571	-20665.91	-23242.45	-23252.51
AIC value	40535.14	41363.82	46566.90	46609.02
BIC value	40910.72	41475.10	46852.06	46970.69

6.5 CONCLUSIONS

This chapter formulates a latent class-based driving behaviour framework that considers drivers' intents for modelling vehicles' 2D movements while recognizing the MVA effect on these movements in HD traffic conditions. Specifically, five extensions are proposed to a typical stimulus-response based driving behaviour framework. First, the subject vehicle's two-dimensional movements are represented as a combination of the angular direction of movement with respect to the longitudinal direction and the magnitude of acceleration or deceleration along the angle. Second, a latent class framework is used to model the driver's intents (latent to the analyst) in two dimensions: (a) the intent to accelerate, decelerate, or maintain a constant speed, and (b) the intent to steer left, right, or straight with respect to the longitudinal axis.

Third, the MVA effect is accommodated to recognize that drivers consider stimuli from multiple vehicles in their vicinity. Fourth, a multi-stimuli model of acceleration is formulated based on the assumption that drivers choose an angular direction of movement that allows them to move with the highest (lowest) possible longitudinal acceleration (deceleration) if they intend to accelerate (decelerate). Fifth, drivers' execution errors are modelled as the difference between their planned acceleration and executed acceleration.

The proposed framework analyses the driver's intents and subject vehicles' two-dimensional movements using a two-stage model system that comprises the following components: (a) a latent class component for analysing drivers' higher-level decisions or intents, and (b) intent-specific model components (class-specific models) for analysing drivers' lower-level decisions such as the specific steering angular direction and the extent of acceleration or deceleration. Properties of the multivariate normal distribution are employed to derive the likelihood function for such a model system. An empirical application of the proposed framework is presented for analysing driver behaviour of motorised two-wheelers using an HD traffic conditions trajectory dataset from Chennai, India.

The empirical results suggest that a driver's higher-level decisions (intents) are not only affected by the immediate lead vehicle but also by vehicles on the left front, right front, left side, and right side of the driver's vehicle, indicating the importance of considering the MVA effect. The findings also show that the microscopic traffic environment factors have a greater impact on drivers' higher-level intents than on their lower-level decisions. Probably, drivers invest greater cognitive resources in making their higher-level, strategic intents than what they invest in making the lower-level, tactical decisions. Furthermore, the empirical results offer insights into the driving behaviour observed in Indian traffic streams. For example, when drivers have an opportunity to pass the slow-moving lead vehicle, they often do so on the right side.

CHAPTER 7 MACROSCOPIC TRAFFIC FLOW PROPERTIES OF THE DRIVER BEHAVIOUR MODELS DEVELOPED IN THIS DISSERTATION

Abstract

In this chapter, we examine the driver behaviour models developed in this dissertation for their ability to reflect the typically observed fundamental relationships between macroscopic traffic flow variables (i.e., flow, density, and space mean speed of traffic streams). To do so, using the proposed driver behaviour models, we develop a traffic simulator for simulating heterogeneous, disorderly (HD) traffic streams. Specifically, we focus on simulating the vehicular movements of cars and motorised two-wheelers. The longitudinal movements of cars are simulated using the discrete-continuous multi-vehicle anticipation model developed in Chapter 3. In addition, the two-dimensional, multi-vehicle anticipation-based latent class framework developed in Chapter 6 is employed to simulate the two-dimensional movement of motorised two-wheelers. Next, Edie's generalised definitions are applied to the simulated trajectories of the vehicles to estimate the speed, flow, and density of the resulting traffic streams. Subsequently, the pairwise relationships between these variables are examined. It is observed from this analysis that the patterns in the pairwise relationships from the simulated data are similar to those that are typically observed in the real world. For instance, low-density values are related to higher speeds and lower flow rates. Similarly, it is observed that when density increases, flow increases up to a maximum value and then decreases. Furthermore, we do not observe inconsistencies in the typical speed–density, speed–flow, and flow–density relationships. All these observations suggest that the developed driver behaviour models in this dissertation are able to reflect the typically observed relationships between macroscopic traffic flow variables.

7.1 INTRODUCTION

The previous chapters focused on the development of driver behaviour models. In this chapter, the developed models are examined for their capability to depict the macroscopic properties of traffic flow.

The macroscopic approach describes the aggregate behaviour of traffic streams using variables such as flow, density, and space mean speed of the streams. The macroscopic properties of traffic flow can be expressed as relationships between these variables. For example, the Greenshields model's assumption of a linear relation between speed and density is a macroscopic property. The variables flow (q), density (k), and space mean speed (u) are called the fundamental traffic flow variables since they are the primary macroscopic descriptors of traffic flow. The following fundamental relationship between these variables is used in combination with an assumed relation between speed and density.

$$q = k \times u \tag{7.1}$$

Furthermore, fundamental diagrams are used to describe the relationship between these variables. Figure 7.1, for example, shows fundamental diagrams of speed–density, speed–flow, and flow–density reported by Chakroborty and Das (2017) using empirical data from an uninterrupted traffic stream on an urban road section. Each plot in Figure 7.1 depicts a pairwise relationship between speed, flow, and density. Specifically, the top-left plot suggests that speed decreases as density increases. The intercept on the y-axis (i.e., speed axis) represents a low-density situation where the driver may drive at the desired speed without being blocked by slow-moving vehicles. On the other hand, the intercept on the x-axis (i.e., density axis) represents the stop-and-go condition scenario where the road is jammed, and the speeds of the vehicles are near zero. The top-right plot between flow and speed suggests that the flow is zero either because the road is nearly empty with a few vehicles moving at free-flow speed, or the road is jammed so that vehicles present on the road cannot move. The bottom-left plot shows a relationship between flow and density. A parabolic curve may be used to represent this relationship when the speed-density relationship is linear. As expected, starting from the origin (i.e., $k = 0$, $q = 0$), flow increases as density increases. This trend continues until the flow peaks. After this point, the flow begins to drop as density continues to increase, and the flow becomes zero when the density reaches jam density. Note that apart from the parabolic shaped fundamental diagram presented in the bottom-left of Figure 7.1, the triangular Newell-Daganzo flux function fundamental diagram (Gordon F Newell, 1993; Daganzo, 1995) is also widely studied in the literature. This flux function is a piecewise linear function of the density with different slopes in the free-flow and congestion regions.

Among the three pairwise relationships between flow, density, and speed, the essential relationship is between speed and density, and the other relationships are implied (Chakroborty and Das, 2017). Once speed and density are related, the other relationships become implied because of the fundamental relation of traffic flow described in Eq. (7.1).

A realistic driver behaviour model should be capable of reflecting the macroscopic properties of traffic flow (Brackstone and McDonald, 1999). Therefore, newly developed driver behaviour models are often tested for their capability to reflect the typically observed fundamental relationships between macroscopic traffic flow variables (flow, density, and space mean speed). The natural next question is how to obtain the macroscopic traffic flow variables using a microscopic driver behaviour model? The answer is by aggregating the individual microscopic driver behaviour. The aggregation can be done analytically or by simulating vehicle trajectories. Here, we adopt the simulation approach since it provides more flexibility and ease in developing the road environment (for example, geometric features of the roadway section, installing traffic signals, etc.) for which the fundamental diagrams are obtained, thereby assisting in examining the model in a wide variety of scenarios. Therefore, using the proposed driver behaviour models, we developed a traffic simulator for simulating *heterogeneous, disorderly* (HD) traffic streams. Another reason for using simulation is that the proposed driver behaviour models may not lend themselves to the analytical exploration of the macroscopic traffic flow properties.

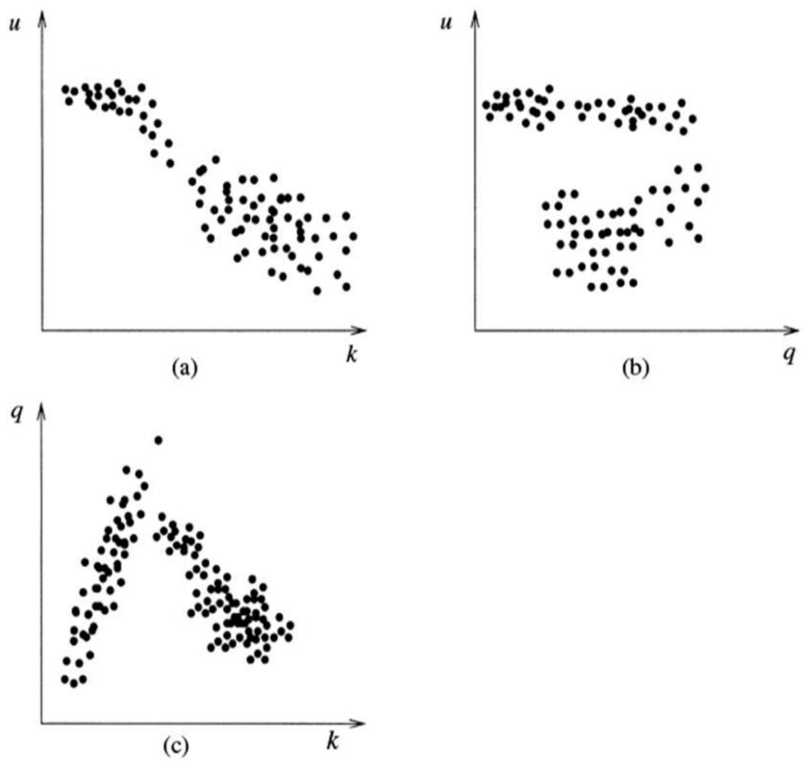


Figure 7.1 Typical plots of (a) speed–density, (b) speed–flow, and (c) flow–density relationships for an uninterrupted traffic stream (Source: Chakroborty and Das, 2017)

The remainder of the chapter is organised in the following manner. Section 7.2 describes the traffic simulator developed to simulate HD traffic streams. Section 7.3 presents macroscopic traffic flow properties of the driver behaviour models developed in this dissertation.

7.2 DESCRIPTION OF THE DEVELOPED TRAFFIC SIMULATOR

MATLAB – a multi-paradigm programming language – is employed for building the simulator in this dissertation. The developed simulator has three components – 1) agents, 2) traffic environment, and 3) user interface. The following subsections go over each of the components and their specifications that were used in this study.

7.2.1 Agents

Autonomous decision-making entities in agent-based modelling are called agents (Bonabeau, 2002). Each agent individually assesses its situation and makes decisions based on a set of rules appropriate for the system they are a part of. This study views different vehicles and traffic stream as agents and system, respectively. The traffic simulator developed in this

consists of two types of agents – cars and motorised two-wheelers. The specifications (physical and operational characteristics) of these agents are provided in Table 7.1.

Table 7.1 Parameters for cars and motorised two-wheelers

Parameter	Cars	Motorised two-wheelers
Dimension	length = 4.2 m	length = 1.8 m
	width = 1.7 m	width = 0.6 m
Desired speed (assumed)	16.67 m/s i.e., 60.00 km/hr	16.67 m/s i.e., 60.00 km/hr

To simulate the movements of agents, we used driver behaviour models developed in this dissertation (Chapter 3 and Chapter 6). The procedures for simulating the movements of agents are outlined below.

7.2.1.1 Simulation of cars' movements

The longitudinal movement of cars is simulated using the model proposed in Chapter 3. The following procedure was followed.

1. Simulate the discrete decisions: To simulate discrete decisions of acceleration, deceleration, or maintain same speed; we used a three-step approach:

   (1a) Calculate the probability of each possible discrete choice decision ($i = a, d, s$) using the following multinomial logit expression.

$$P_{qi} = \frac{\exp\left(\beta_i^T x_{qi}\right)}{\sum_j \exp\left(\beta_j^T x_{qj}\right)};\ i = a, d, s \tag{7.2}$$

   (1b) Use the above calculated discrete choice probabilities to simulate the discrete decision of acceleration ($i = a$), deceleration ($i = d$), or maintain constant speed $(i = s)$. This can be done by drawing a pseudo-random number from a uniform [0,1] distribution and seeing where the draw falls in the cumulative probability space of the discrete choices $(i = a, d, s)$.

2. For a given simulated discrete decision of acceleration or deceleration ($i = a, d$), simulate the corresponding continuous decision (acceleration and deceleration) values. To do so, compute the conditional expected value of the corresponding continuous decision (i.e., extent of acceleration or extent of deceleration) using the following expression:

$$E\left[f\left(m_{qi}|i\right)\right]=\int_{L_i}^{U_i}\left(f\left(m_{qi}|i\right)m_{qi}dm_{qi}\right);\ i=a,d \tag{7.3}$$

where, $f\left(m_{qi}|i\right)$ is the conditional density of the corresponding continuous decision (acceleration and deceleration) and takes the following expression, and all other terms are defined in Chapter 3:

$$f\left(m_{qi}|i\right)=\left(P_{qi}\right)^{-1}\left\{\frac{1}{\sigma_{\eta i}}f_{\eta i}\left(\frac{m_{qi}-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)-\frac{1}{\sigma_{\eta i}}f_{\eta i}\left(\frac{m_{qi}-\alpha_i^T z_{qi}}{\sigma_{\eta i}}\right)\frac{\partial C_\theta\left(u_{q1}^i,u_{q2}^i\right)}{\partial u_2^i}\right\};\ i=a,d \tag{7.4}$$

3. Update the longitudinal position of the vehicle using the acceleration as $E\left[f\left(m_{qi}|i\right)\right]$ and specified update time.

7.2.1.2 Simulation of motorised two-wheelers' movements

The model proposed in Chapter 6 is used to simulate the two-dimensional movement of two-wheelers. The following procedure was followed to do so.

1. Simulate driver's intents latent to the analyst: Calculate the probability of each possible intent using the following logit expression.

$$P_{qts}=\frac{\exp\left(\beta^T x_{qts}\right)}{\sum_s \exp\left(\beta^T x_{qts}\right)};\ s=AL,AS,AR,CL,CS,CR,DL,DS,DR \tag{7.5}$$

   Use the above calculated probabilities to simulate the intents. This can be done by drawing a pseudo-random number from a uniform [0,1] distribution and seeing where the draw falls in the cumulative probability space of the latent intents.

2. Simulate radial cone choice: For a given simulated latent intent, simulate the corresponding latent-intent specific radial cone choice. To perform this, predict the acceleration or deceleration on each radial cone using the following expression:

$$A_{q(t+\Delta t)i}^s=\cos\left(\theta_i\right)f\left(\Delta AD_{qti}\right)\left[\alpha_{s1}+\alpha_{s2}\left(V^{des}-V_{qt}\right)+\alpha_{s3}\left[\delta_i\left(\Delta X_{qti}-\Delta X_{qti}^{des}\right)+\left(1-\delta_i\right)\left(\Delta X_{qti}^{curb}-\Delta X_{qti}^{des}\right)\right]+\alpha_{s4}\delta_i\Delta V_{qti}\right]+\varepsilon_{qti}^s \tag{7.6}$$

   where, ε_{qti}^s are normally distributed error terms with zero mean and unit variance. The radial cone choice would be the cone that gives $\max\left(A_{q(t+\Delta t)i}^s\right)$.

3. To simulate the continuous decision (extent of acceleration or deceleration): $Max\left(A_{q(t+\Delta t)i}^s\right)$ is the driver's intended acceleration value. However, as discussed in

Chapter 6, the actual executed acceleration value is likely to differ from what the driver intended due to vehicle capabilities, drivers' driving expertise, machine error, and other unobserved factors and this difference is labelled as the execution error. Hence, to calculate the actual extent of acceleration or deceleration, simulate a normally distributed error term (will be used as an execution error) with zero mean and σ_s^2 variance and add it to $\max\left(A^s_{q(t+\Delta t)i}\right)$.

4. Update the position of the vehicle using the predicted radial cone and the computed extent of acceleration/deceleration for the specified update time.

7.2.2 Environment

A traffic simulator environment is a virtual world where the agents are created, run and displayed. The user defines the length of the road and the number of lanes based on the different traffic scenarios being examined. In this study, the width of each lane is fixed to 3.5 m, and only one-way movements are simulated. Different congestion levels on the road can be simulated by placing a traffic signal at the end of the road section with cycle length as input from the user. Further, the user can set the simulation duration.

Note that the traffic environment used in this study is simplistic in nature. Specifically, a one-way midblock road section is considered. As the driver behaviour models developed in this dissertation are for the one-way midblock road section, this simplistic traffic environment is sufficient to test whether the developed models are capable of reproducing the fundamental diagram. A more complex traffic environment can be simulated in future work.

7.2.2.1 Vehicle generation

Vehicles are generated one after the other following a time-headway. The headways are drawn from a normal distribution, $N\left(TotalSimulationTime/(SimTime+100), 0.5^2\right)$ s following the study by Lee (2007). Here, the mean headway decreases as the simulation time $(SimTime)$ increases. This vehicle generation mechanism helps to simulate the traffic that gradually changes from free flow to congested flow. Notably, a normal distribution leaves scope for negative headways. However, the probability of generating such negative headways is very small for the assumptions we made. Hence, we observed very few instances (that too at large simulation times) where negative headways were generated. In such cases, we assumed the headway as 0.5 s.

Note that all vehicles are generated at the start-point of the road section. The type of vehicle is randomly assigned using sa random number generated based on the assumed traffic composition. Once the vehicle is generated, its physical characteristics (e.g., length and width of the vehicle) and operational characteristics (e.g., free-flow speed) are assigned based on its vehicle type (according to Table 7.1).

7.2.2.2 Vehicle placement at the start of the road section

The following procedure is followed for vehicle placement while simulating HD traffic streams.

1. Vehicle's lateral position: Since HD traffic streams are simulated, there is no concept of lane discipline, and the generated vehicle can take any lateral position on the road. Hence, the lateral position for the generated vehicle is assigned randomly within the width of the road using a uniform distribution. Note that every generated vehicle scans for spaces throughout the entire width of the road. During the scanning, leaders are identified. If leaders are present, the availability of safe gaps is checked. Here, the gap is calculated as the longitudinal distance between the generated vehicle and its immediate lead vehicle and checked against a safe gap as the length of the vehicle. If this calculated gap is less than the safe gap, the generated vehicle is shifted laterally until a safe gap becomes available. Once this condition is satisfied, the vehicle is placed laterally. If not, the generated vehicle is removed from the queue.
2. Vehicle's longitudinal position: The generated vehicles are placed at the beginning of the road stretch with randomly assigned speeds that follow a normal distribution with 3 m/s mean and unit variance.

7.2.2.3 Signal control

A traffic signal is installed near the end of the road section. The cycle length of the signal is manipulated by the user depending on the traffic scenario being tested.

7.2.3 User Interface

The user interface allows the traffic simulator to receive commands and instructions from users. It also helps with the collection of data from agents. As shown in Figure 7.2, the traffic simulator developed in this dissertation has a graphical user interface (GUI). The GUI is divided into the main window, control panel, and simulation window. The main window depicts the simulation's spatial organisation. Users can adjust the simulator's attributes and parameters using a control panel. The simulation window allows one to observe the vehicular movements and the traffic flow.

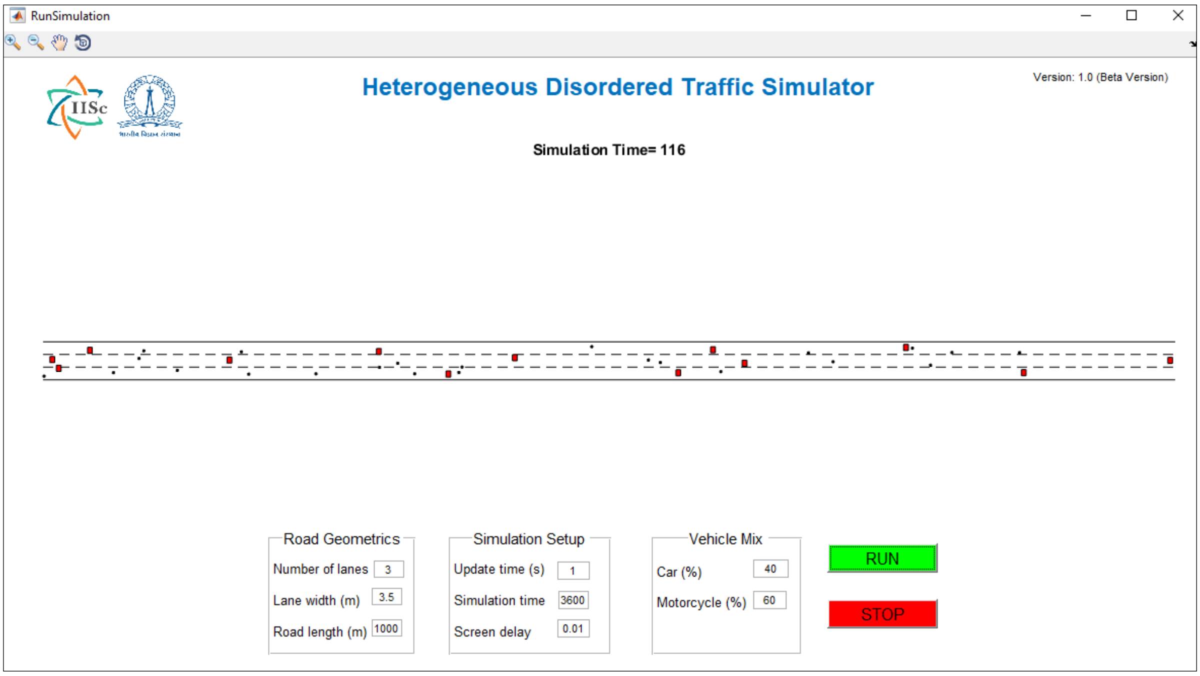


Figure 7.2 Graphical user interface of traffic simulator developed in the dissertation

7.3 MACROSCOPIC TRAFFIC FLOW PROPERTIES OF THE DRIVER BEHAVIOUR MODELS DEVELOPED IN THIS DISSERTATION

Fundamental diagrams of traffic flow are plotted after aggregating the simulated trajectories. A vehicle trajectory describes the spatial position of the vehicle over time along the roadway. Using the simulator, the analyst first assumes a hypothetical roadway section with its geometrical properties, sets the parameters of the driver behaviour model under examination, simulates the trajectories of the vehicles using the model, and finally, measures all relevant macroscopic traffic flow quantities. There are different ways to measure flow, density, and space mean speed, such as the highway capacity manual (HCM) method, the n-t method, and the x-t method based on Edie's generalised definition (D. Ni, 2007; Sharma et al., 2021). The HCM method is based on the definitions of flow, density, and speed and their fundamental relationship. The n-t method is based on the cumulative number of vehicles in the cumulative number-time (n-t) domain. In contrast, the x-t method estimates the fundamental variables using trajectory data. The paper by Ni (2007) discusses each of these methods in detail. Edie's generalised definitions are applied when trajectory data is available to calculate the fundamental traffic flow parameters (Edie, 1963; Sharma et al., 2021). Hence, the flow, density, and speed are estimated using Edie's generalised definitions from simulated trajectories in this study. This section discusses Edie's procedure in detail.

7.3.1 Simulation Setup

The driver behaviour models developed in this dissertation were used to generate fundamental diagrams for three specific scenarios. Table 7.2 summarises the setting of each of these scenarios.

Table 7.2 Simulation set up for different scenarios

Parameter	**Scenario 1**	**Scenario 2**	**Scenario 3**
Vehicle type proportion	All vehicles are cars	All vehicles are motorised two-wheelers	Cars = 40 % Motorised two-wheelers = 60 %
Driver behaviour model	Discrete-continuous multi-vehicle anticipation model developed in Chapter 3	Two-dimensional, multi-vehicle anticipation-based latent class framework developed in Chapter 6	Discrete-continuous multi-vehicle anticipation model for cars and two-dimensional, multi-vehicle anticipation-based latent class framework for motorised two-wheelers
Length of the road	1000 m	1000 m	1000 m
Simulation time $(TotalSimTime)$	1800 s	1800 s	1800 s
Vehicle generation	The vehicles are generated with time headways drawn from a normal distribution as follows: $N\left(\frac{TotalSimulationTime}{SimTime+100}, 0.5^2\right)$. The mean of this normal distribution decreases with increasing simulation time $(SimTime)$		
Signal control	Three signal cycles are scheduled. The first signal cycle starts at 500 s and ends at 600 s. The second signal cycle starts at 1000 s and ends at 1100 s. The third signal cycle starts at 1500 s and ends at 1600 s. This signal schedule ensures that the traffic simulation can simulate all trends of traffic stream densities.		

Scenario 1 is used to analyse the driver behaviour model developed for car drivers, whereas Scenario 2 is used to analyse the driver behaviour model developed for motorised two-wheelers. In Scenario 3, a *heterogeneous traffic* stream is simulated that consists of cars and motorised two-wheelers using driver behaviour models built for both car and motorised two-wheeler drivers. In all three scenarios, the length of the road, simulation time and update time are considered as 1 KM, 30 minutes, and 1 s, respectively. Table 7.2 also provides the time headway distribution used to generate vehicles in all three scenarios. As discussed earlier, a traffic signal is installed near the end of the road section. In all scenarios, three signal cycles are scheduled to ensure that the traffic simulation can simulate a wide range of traffic stream densities.

7.3.2 Measuring Macroscopic Traffic Flow Variables from Trajectories

The fundamental traffic flow parameters are estimated from simulated trajectories using Edie's generalised definitions, which use the total time taken, and the distance travelled by each vehicle in the longitudinal position-time region for density and flow estimation. The longitudinal position-time diagram provided in Figure 7.3 demonstrates Edie's method.

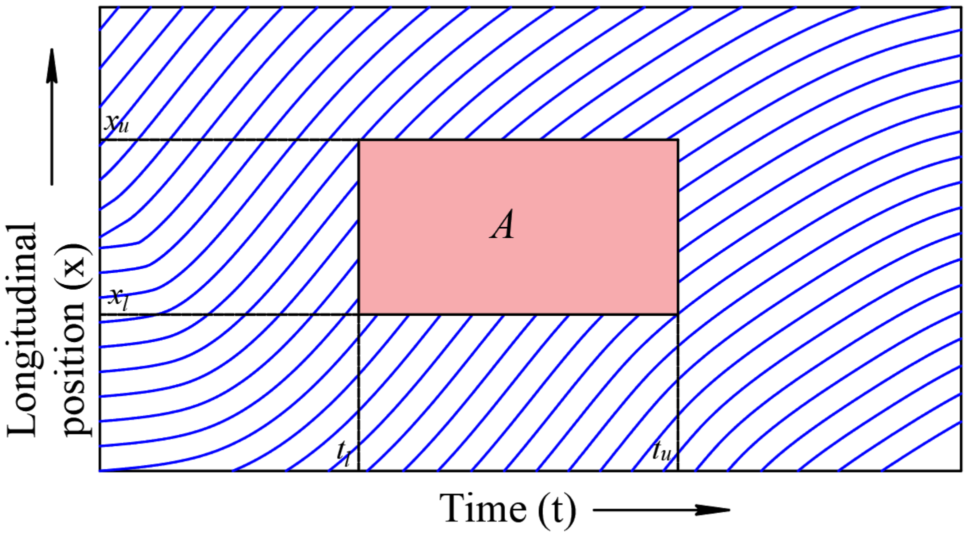


Figure 7.3 Longitudinal Position-Time diagram for vehicle trajectories

According to Edie, density is defined as the total time spent by all vehicles in a region A divided by the area of A (refer Eq. (7.7)). Flow is defined as the total distance travelled by all vehicles in a region A divided by the area of A (refer Eq. (7.8)). Finally, speed is calculated as the ratio of the total distance travelled by all vehicles in A over the total time spent by all vehicles in A (refer Eq. (7.9)).

$$Density = \frac{\sum_{q=1}^{Q} TT_q}{Area} \tag{7.7}$$

$$Flow = \frac{\sum_{q=1}^{Q} d_q}{Area} \tag{7.8}$$

$$Speed = \frac{\sum_{q=1}^{Q} d_q}{\sum_{q=1}^{Q} TT_q} \tag{7.9}$$

where, Q represents the total number of vehicles. TT_q represents the time taken by a vehicle q over a defined rectangular region A. d_q represents the longitudinal distance travelled by vehicle q over the rectangular region A. The quantities TT_q, d_q, and $Area$ are provided below.

$$d_q = \min\left(x^{(q)}(t_u), x_u\right) - \max\left(x^{(q)}(t_l), x_l\right) \tag{7.10}$$

$$TT_q = \min\left(t^{(q)}(x_u), t_u\right) - \max\left(t^{(q)}(x_l), t_l\right) \tag{7.11}$$

$$Area = \left(x_u - x_l\right) \times \left(t_u - t_l\right) \tag{7.12}$$

where, x_l and x_u are the lower and upper bounds of A in X domain, respectively. t_l and t_u are the lower and upper bounds of A in t domain, respectively. $x^{(q)}(t_u)$ and $x^{(q)}(t_l)$ are the locations of the q^{th} vehicle's passes at times t_l and t_u, respectively. $t^{(q)}(x_u)$ and $t^{(q)}(x_l)$ are the time instances when the q^{th} vehicle passes locations x_l and x_u, respectively. In this dissertation, the longitudinal position-time region for simulated trajectories is divided into windows of 20 s (along the time axis) and 100 m to 1000 m (along the longitudinal position axis), i.e., $t_u - t_l = 20\,s$ and $x_u - x_l = 1000 - 100 = 900\,m$. The first 100 m and 80 s of simulation are discarded as a warm-up section and a warm-up period, respectively.

7.1.1 Results and Discussion

The trajectories are generated for each scenario using the developed simulator. Next, Edie's generalised definitions are applied to estimate the speed, flow, and density. Then, the pairwise relationship between speed, flow and density are plotted and provided in Figure 7.4-7.6.

Fundamental traffic flow diagrams are developed for all three scenarios. The patterns in these diagrams are similar to those that are typically observed in the real world (as in Figure 7.1). Also, similar patterns were reported by Thankappan et al. (2010) and Thankappan and Vanajakshi (2015) for Indian traffic conditions. Additionally, we observe that the maximum speed from the simulation results is around 60 km/hr, which is the desired speed set for the agents in the current traffic simulator. Furthermore, we do not observe inconsistencies in the typical density-speed, flow-speed, and density-flow relationships in all three scenarios. For instance, low-density values are related to higher speed and lower flow rates, as expected. Similarly, we observe that when density increases, flow increases up to a maximum and then decreases. All these observations suggest that the driver behaviour models developed in this dissertation are able to reflect expected macroscopic traffic stream relationships. These observations provide some credence to the models developed in this dissertation. Note that this dissertation does not perform stability analysis using the proposed models, which is an important avenue for future research.

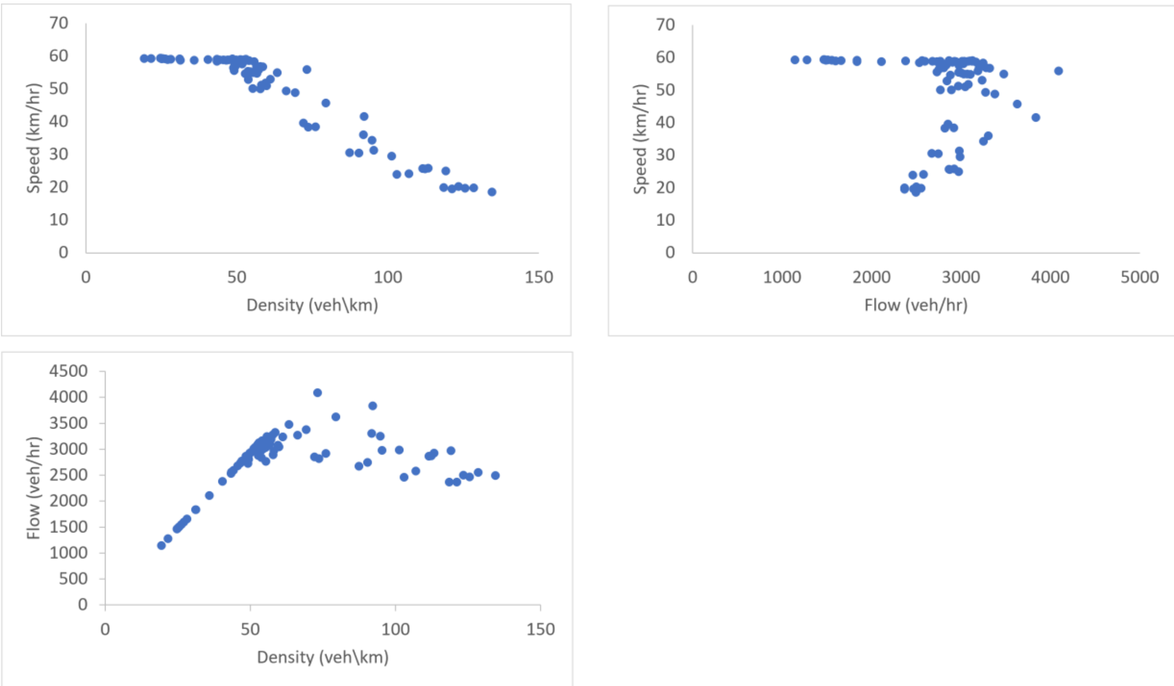


Figure 7.4 Fundamental diagrams for Scenario 1

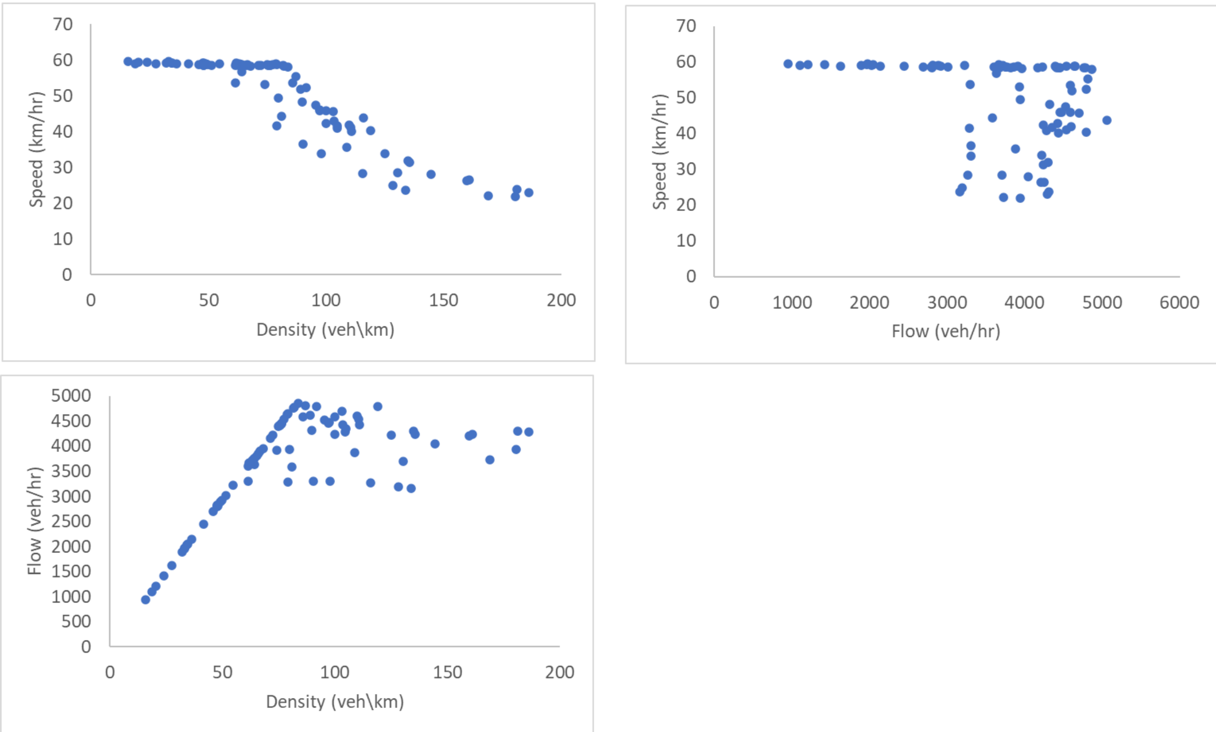


Figure 7.5 Fundamental diagrams for Scenario 2

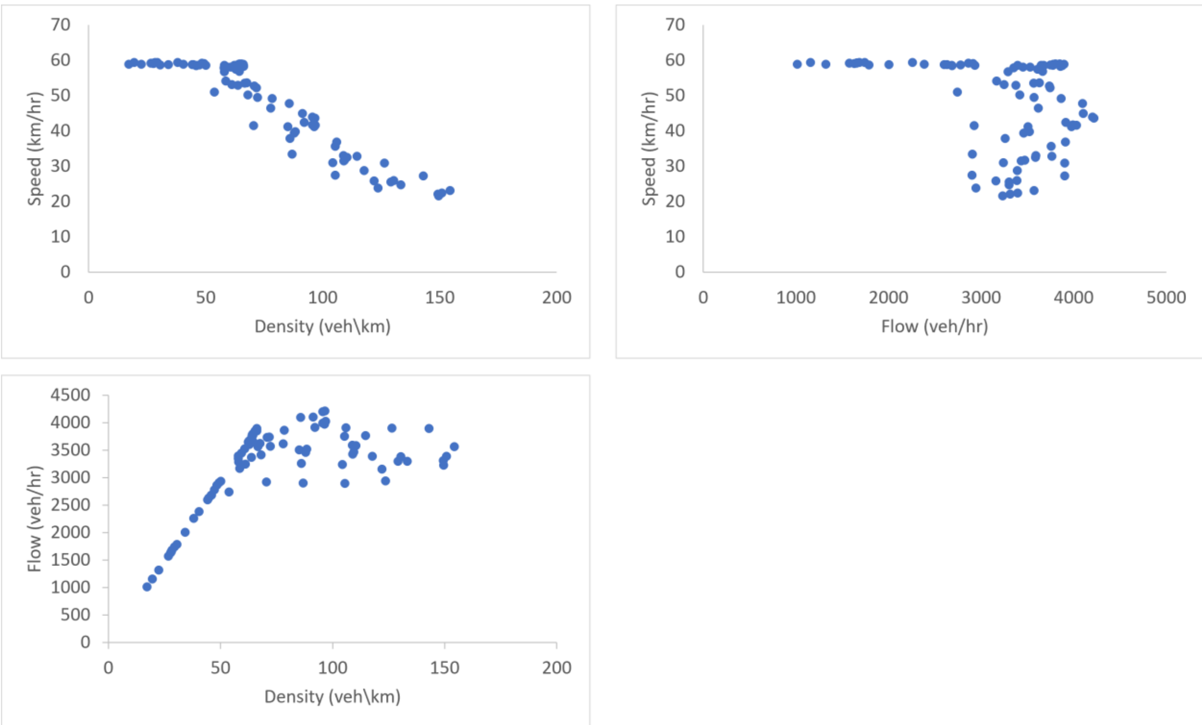


Figure 7.6 Fundamental diagrams for Scenario 3

CHAPTER 8 CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

The last chapter of this thesis re-states the proposed objectives, and the associated research questions, recapitulates the findings from chapters 2 to 7 and highlights the research contributions from two perspectives: methodological and empirical. Finally, it presents the limitations of this thesis and identifies future research directions.

8.1 RESEARCH OVERVIEW

Driver behaviour models are widely used in traffic engineering literature and practice for understanding and describing drivers' manoeuvring decisions in vehicular traffic streams. They also form the building blocks of traffic microsimulation tools, which are used for traffic flow analysis, traffic safety analysis, traffic emission estimation, traffic control studies, etc. Most of these models in the literature are developed for *homogeneous traffic* conditions (typically observed in countries such as Australia, the United States, Germany, and the Netherlands). However, the *heterogeneous, disorderly* (HD) traffic conditions observed in the urban areas of countries like India differ significantly from *homogeneous traffic* conditions. More specifically, in contrast to *homogeneous traffic* conditions, HD traffic streams include a wide range of vehicle classes with varying physical and operational characteristics, and vehicles demonstrate weak to no lane discipline and a large extent of lateral movements.

A review of the literature on driver behaviour models (in Chapter 2) suggests that there has been a substantial amount of previous research on modelling driver behaviour in *homogeneous traffic* streams. There has also been an increasing interest in modelling driver behaviour in HD traffic streams. However, there are still several research gaps in this area, as identified below (and discussed in detail in earlier chapters):

- Inadequate consideration of the multi-vehicle anticipation (MVA) effect while modelling driver behaviour in HD traffic streams,
- Limited efforts to consider driver behaviour as a combination of different manoeuvring decisions, such as the decision of whether to accelerate, decelerate, or remain in same speed (represented as a discrete variable) and the decision of the extent of acceleration or deceleration (represented as continuous variables) – as opposed to using a single, continuous variable to represent all these facets of driver behaviour,
- Inadequate attention to (and lack of methods for) modelling drivers' perception errors in MVA-based driver behaviour models, and

- Limited attention to incorporating two-dimensional (2D) movement of vehicles in HD traffic streams while also considering drivers' intentions (that are latent to the analyst) and the MVA effect.

In view of the above gaps in the literature, this dissertation set out to develop driver behaviour models for HD traffic streams on uninterrupted traffic facilities while considering the following aspects – (1) the MVA behaviour, where drivers' manoeuvring decisions are influenced by multiple vehicles around them, as opposed to a single lead vehicle ahead (2) the treatment of driver behaviour as a combination of different manoeuvring decisions, such as the decision of whether to accelerate, decelerate, or remain in same speed and the decision of the extent of acceleration or deceleration (as opposed to a single, continuous variable representing the driver behaviour), (3) the incorporation of stochasticity due to drivers' perception errors, and (4) the consideration of 2D movements and driver's intentions (latent to the analyst) simultaneously while also incorporating MVA behaviour. The specific methodological and substantive objectives of the dissertation are mentioned below.

Methodological objectives (M1 to M4)

M1: To develop a discrete-continuous choice modelling framework for describing car drivers' longitudinal movements in HD traffic conditions. This objective serves the following two purposes: (1) the consideration of the MVA effect while modelling driver behaviour in HD traffic streams, and (2) the consideration of the driver's discrete decisions (i.e., the decision of whether to accelerate, decelerate, or maintain a constant speed, represented as a discrete variable) and continuous decisions (i.e., the extents of acceleration and deceleration, represented as continuous variables) separately but also model them simultaneously.

M2: To enhance the above-mentioned discrete-continuous modelling framework to recognise subject vehicle- and driver-specific unobserved factors that influence driver behaviour.

M3: To incorporate drivers' perception errors for variables describing the traffic environment in discrete choice-based models of driver behaviour. And to evaluate two different ways of specifying stochasticity due to drivers' errors in their perception of the traffic environment – additive stochasticity and multiplicative stochasticity.

M4: To develop an MVA-based latent class framework to simultaneously model 2D movements of motorised two-wheelers. And to develop a latent class framework to incorporate the MVA effect and the driver's intentions (those are latent to the analyst) along two dimensions – (a) the intent to accelerate, decelerate, or maintain a constant speed, and (b) the intent to steer to the left of, right of, or straight along the longitudinal direction.

Substantive objectives (S1 to S5)

S1: To demonstrate the importance of MVA using the model formulated for objective **M1** in two different empirical settings – (1) an HD traffic stream setting using a trajectory dataset from the city of Chennai, India and (2) a *homogeneous traffic* stream setting using a trajectory dataset from the United States of America (USA). Another objective is to compare and contrast car driver behaviour between HD and *homogeneous traffic* conditions using these two trajectory datasets.

S2: To apply the model formulated for objective **M1** to test whether the factors influencing the decision to accelerate or decelerate are different (or have a different influence on) than the factors influencing the extent of acceleration or deceleration.

S3: To apply the model formulated for objective **M3** to evaluate the importance of incorporating drivers' perception errors in traffic environment variables vis-à-vis allowing unobserved heterogeneity in drivers' response to those variables in driver behaviour models.

S4: To apply the model formulated for objective **M4** to test whether the extent of cognitive efforts invested for making higher-level decisions (drivers' intent to accelerate, decelerate, or maintain a constant speed, and intent to steer to the left of, right of, or straight along the longitudinal direction) are different from those invested for lower-level decisions (decisions of exactly how much to accelerate or decelerate and which specific angular direction to move along).

S5: To evaluate the driver behaviour models developed in this dissertation for their ability to mimic the macroscopic traffic flow properties of uninterrupted traffic streams observed in HD traffic conditions. To achieve this, another objective is to develop a traffic simulator for simulating HD traffic streams using the proposed driver behaviour models.

8.2 METHODOLOGICAL CONTRIBUTIONS

8.2.1 An MVA-Based Discrete-Continuous Choice Modelling Framework to Model Car Driver Behaviour in HD Traffic Streams

Chapter 3 formulated an MVA-based discrete-continuous choice modelling framework for describing car driver behaviour in HD traffic conditions. To incorporate MVA, the concept of an *influence zone* around a vehicle (subject vehicle) was introduced, and it was assumed that vehicles within the *influence zone* influence the subject vehicle's driving behaviour. Further, driving decisions were characterised as a combination of discrete and continuous components. The discrete component involved the decision to accelerate, decelerate, or maintain a constant speed and the continuous component involved the decision of how much to accelerate or decelerate. A copula-based joint modelling framework that allows dependencies between discrete and continuous components was proposed. Such a joint modelling framework recognises that the discrete and continuous decisions are made simultaneously, and common unobserved factors influence both decisions. Additionally, truncated distributions were employed for the continuous model components to avoid the prediction of unrealistically high acceleration or deceleration values.

8.2.2 A Panel Data-Based Discrete-Continuous Choice Modelling Framework to Analyse Longitudinal Driver Behaviour in Homogeneous and HD Traffic Conditions

In Chapter 4, we proposed a panel data-based discrete-continuous choice modelling framework to analyse car driver behaviour in two disparate trajectory datasets – one from an HD traffic stream in India and another from a *homogeneous traffic* stream in the USA. This panel data-based model was built on an MVA-based discrete-continuous choice model proposed in Chapter 3. The panel data-based framework allows the analyst to isolate the subject vehicle- and driver-specific unobserved factors (such as age and aggressiveness) that do not vary across different observations of the same vehicle but have an influence on the vehicle's driver behaviour. Doing so helps reduce the confounding effects of such unobserved factors on analysing the influence of observed factors (such as relative speeds and spacing between the subject vehicle and other vehicles) on driver behaviour. This also helps in reducing the confounding effects of unobserved factors on analysing the differences in driving behaviour between HD and *homogeneous traffic* streams.

8.2.3 Mixed Multinomial Logit Based Framework to Consider Drivers' Perception Errors in Driver Behaviour Models

Chapter 5 presented a mixed multinomial logit based framework to represent driver's perception errors in discrete choice models of driver behaviour. This framework helped us examine the importance of incorporating perception errors in driver behaviour models, the repercussions of ignoring such errors, and an understanding of which variables are perceived with a greater uncertainty/error.

Econometric analysis was undertaken to evaluate two different ways of specifying perception errors in the choice environment variables in discrete choice models – (a) additive specification and (b) multiplicative specification. It was shown that models with an additive error specification run into parameter identification issues when the analyst attempts to accommodate perceptions errors in a large number of traffic environment variables. On the other hand, models with a multiplicative error specification are not, in theory, saddled with such parameter identification problems. The usefulness of the proposed framework with multiplicative errors is demonstrated through simulation experiments as well as an empirical application for analysing driver behaviour while considering drivers' errors in perceiving traffic environment variables. Empirical results from Chapter 5 suggested 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.

8.2.4 A Two-Dimensional, Multi-Vehicle Anticipation, and Multi-Stimuli Based Latent Class Framework to Model Driver Behaviour in HD Traffic Streams

As discussed in Chapter 6, driving manoeuvres in HD traffic streams involve multifaceted decisions such as: (a) the intention to accelerate, decelerate, or maintain a constant speed, (b) the extent of intended acceleration or deceleration, (c) the intention to steer to the left of, right of, or straight along the traffic flow direction, and (d) the specific angular direction of movement. These decisions must be made quickly based on the driver's perceptions of the constantly evolving traffic environment around them. However, from a cognitive science standpoint, humans are endowed with a limited amount of cognitive resources such as working memory they need to store and process information for making their decisions (Sweller, 1988).

Therefore, drivers might allocate their cognitive resources optimally to quickly make their manoeuvring decisions. Specifically, given the multifaceted decisions drivers need to make in a short timeframe and the complexity of the traffic environment around them, it is plausible that they break down their decision-making into manageable steps for cognitive ease. In this context, there is scope to explore if analysing drivers' intents (which are latent to the analyst) first, followed by the specific actions they take, can help in filling this gap. For instance, it may be that higher-level, strategic decisions – such as the intentions of whether to accelerate, decelerate, or maintain a constant speed and whether to steer to the left of, right of, or keep straight along the longitudinal direction – are made first, followed by lower-level, tactical decisions – such as exactly how much to accelerate or decelerate and which specific angular direction to move along. In this context, we conjecture that a greater amount of cognitive effort might be invested in making the higher-level, strategic decisions or intentions than that in making lower-level, tactical decisions. This is likely because a greater amount of information is processed for making higher-level intentions than that for making lower-level decisions. The Chapter 6 aims at gathering evidence toward this conjecture using statistical analysis of the trajectory data, without necessarily delving into measurements of the cognitive loads exerted in making the above-mentioned decisions.

In view of the above discussion, Chapter 6 formulated a latent class-based driving behaviour framework that considers the driver's strategic intentions (those are latent to the analyst) for modelling motorised two-wheeler's 2D movements while considering the MVA effect on these movements in HD traffic conditions. Specifically, the following extensions are proposed to a conventional stimulus-response based driving behaviour framework:

1. The observed 2D movements of a subject vehicle (SV) are represented as a combination of the angular direction (or orientation) of movement with respect to the longitudinal direction and the magnitude of acceleration or deceleration along the direction.
2. The observed 2D movements of an SV at a time instance are assumed to be a result of a sequential decision-making process of its driver, where higher-level (strategic) decisions precede lower-level (tactical) decisions. The higher-level decisions are drivers' intents along the following two dimensions:
    (a) the intent to accelerate, decelerate, or maintain a constant speed, and
    (b) the intent to steer to the left of, right of, or straight along the longitudinal direction.

The lower-level decisions are the tactical decisions of how much to accelerate or decelerate and which specific angular direction to move along.

A latent class framework is utilised to model drivers' intents as they are latent to the analyst. Conditional on the intents, the lower-level decisions are modelled. The sequential decision-making assumption combined with our two-stage modelling framework allows us to gather evidence toward our conjecture that the extent of cognitive effort needed for making higher-level intents is different from those for making lower-level choices. In the absence of measurements of cognitive load, this is achieved by comparing the strength of influence of various traffic environment variables on the higher- and lower-level decisions.

3. The MVA effect is accommodated, wherein the driver of an SV is assumed to consider stimuli from multiple vehicles within its *influence zone* for making their intents and manoeuvring decisions.
4. For the lower-level decisions, a multi-stimuli model of acceleration is formulated based on the assumption that drivers choose a specific angle of movement that allows them to move with the highest (lowest) possible longitudinal acceleration (deceleration) if they intend to accelerate (decelerate).
5. Drivers' execution errors are modelled as the difference between their intended extent of acceleration or deceleration and the executed acceleration values.

Finally, properties of the multivariate normal distribution are employed to derive the likelihood function to estimate the parameters of the proposed model.

8.3 EMPIRICAL FINDINGS

In addition to the above-discussed methodological contributions, the substantive contributions from this dissertation were in the context of analysing driver behaviour models using vehicular trajectory datasets. The following subsections summarise some key findings and contributions from the empirical analyses undertaken in this dissertation.

8.3.1 Importance of Considering MVA in Driver Behaviour Models

The results presented in Chapters 3 and 4 demonstrated the importance of considering MVA for describing driving behaviour in HD traffic conditions as well as *homogeneous traffic* conditions. Specifically, the results lent credence to the idea of an *influence zone* around the vehicle and corroborated our hypothesis that driving behaviour is influenced by multiple vehicles within the *influence zone* of the subject vehicle. Further, drivers in HD traffic conditions not only consider vehicles that are ahead of their vehicle but also consider those vehicles that are on either side. At the same time, while drivers' decision to accelerate and the extent of acceleration is governed by multiple vehicles ahead, their decision on the extent of deceleration is likely to be affected more by the immediate lead vehicle than other vehicles in the *influence zone*. Such nuanced findings have not been reported by other studies that considered MVA in driving behaviour.

8.3.2 Empirical Findings on Driver's Discrete and Continuous Decisions

The results presented in Chapters 3 and 4 supported the notion that the driver behaviour can be represented as a set of discrete and continuous decisions, and the factors influencing the discrete decisions might be different or may affect the continuous decisions differently. Specifically, not all traffic environment variables found to influence the discrete decisions were found influential on continuous decisions (or how much to accelerate or decelerate) and vice versa. Moreover, the influence of several variables was found to be stronger on the decision to accelerate or decelerate than on the decision of how much to accelerate or decelerate. This suggested that the discrete and continuous decisions likely require different cognitive efforts by drivers and are influenced in different ways by the traffic environment variables.

8.3.3 Identification and Characterisation of the Influence Zones for Determining Potential Leaders to Predict Acceleration/Deceleration Behaviour

This dissertation introduced the concept of an *influence zone* which was defined as a hypothetical zone within which the surrounding traffic environment, including vehicles, road boundaries, stationary traffic control devices, etc., influences driver behaviour. Specifically, we assumed a rectangular-shaped *influence zone*. The reasons behind choosing a rectangular-shaped influence zone were as follows: a) road boundaries for straight sections can be approximated as two edges of a rectangular-shaped *influence zone*, and b) its simple geometric properties provide higher computational tractability. For example, one can easily divide the rectangular *influence zone* into different compartments to examine the effect of different vehicles on the subject vehicle. As the vehicle's longitudinal position shifts to the left or right of the roadway width, the *influence zone* gets truncated accordingly in that direction.

After an extensive empirical investigation, this dissertation recommended the appropriate size of an *influence zone* for drivers of HD and *homogeneous traffic* streams. Notably, a 60 m length *influence zone* was found to be more suitable than shorter length zones to model driver behaviour in HD traffic conditions. In contrast, a 30 m length *influence zone* was found to be more suitable for *homogeneous traffic* conditions.

8.3.4 Car Driver Behaviour in HD and Homogeneous Traffic Streams

The empirical results presented in Chapter 4 revealed both similarities and differences in car driver behaviour between *homogeneous traffic* and HD traffic conditions' trajectory data. Specifically, in both traffic conditions' trajectory data, in addition to vehicles ahead of the subject vehicle in its lane, vehicles ahead in the adjacent lanes influence its driver behaviour. However, side vehicles influence drivers' decision-making only in HD traffic conditions. The insights from Chapter 4 can assist in developing behaviourally realistic driver behaviour models specific to homogeneous and HD traffic conditions. Specifically, such considerations and the modelling framework presented in Chapter 4 can potentially help in better simulating traffic flow in HD traffic conditions.

8.3.5 Role of Perception Errors in Driver Behaviour

The newly proposed modelling framework in Chapter 5 provided insights into the role of perception errors in driver behaviour. For example, first, the empirical application demonstrated that allowing for perception errors in traffic environment variables is more important than allowing unobserved heterogeneity in drivers' responses to those variables. Second, 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. Third, 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. Fourth, drivers' perception of lateral gaps between two moving vehicles ahead was associated with greater uncertainty than that associated with longitudinal space gaps with respect to any of those vehicles. Fifth, as expected, it was difficult to recover variability due to perception errors for variables that did not have a significant influence on the choice outcome.

8.3.6 Insights on Driver Behaviour of Motorised Two-Wheelers in HD Traffic Streams

The empirical results from Chapter 6 suggested that a driver's higher-level decisions (intents) are not only affected by the immediate lead vehicle but also by vehicles on the left front, right front, left side, and right side of the driver's vehicle, indicating the importance of considering the MVA effect. The findings also showed that the microscopic traffic environment factors have a greater impact on drivers' higher-level intents than on their lower-level decisions. Probably, drivers invest greater cognitive resources in making their higher-level, strategic intents than what they invest in making the lower-level, tactical decisions. Furthermore, the empirical results offer insights into the driving behaviour observed in Indian traffic streams. For example, when drivers have an opportunity to pass the slow-moving lead vehicle, they often do so on the right side. The proposed framework and the insights from Chapter 6 might inform improvements to existing approaches to simulate motorised two-wheeler movements in non-lane-based HD traffic streams.

8.4 LIMITATIONS AND FUTURE RESEARCH DIRECTIONS

This dissertation provided several methodological advances in modelling driver behaviour in HD traffic conditions. The results provided in this dissertation demonstrated that the developed driver behaviour models are realistic and robust. However, the work presented in this dissertation is not without limitations. The following subsections identify and discuss the potential limitations of the proposed work. In doing so, we also identify some future research possibilities in this research area.

8.4.1 Measurement of Drivers' Cognitive Resource Allocation for Making Complex Driving Decisions

The empirical applications of developed driver behaviour models are used to make informed speculations on how drivers might allocate their cognitive resources for making complex driving decisions. Note that this dissertation does not measure drivers' cognitive efforts directly but compares the strength of influence of traffic environment variables on their driving behaviours to arrive at plausible conclusions on the extent of cognitive resources invested by drivers in making different manoeuvring decisions. In this context, it would be useful to device experiments to measure drivers' cognitive resource allocation as it relates to their driving behaviour.

8.4.2 Heterogeneity in the Size and Shape of the Influence Zone

In this dissertation, we restricted ourselves to a homogeneous and static, rectangular *influence zone*, even though we recognised that the *influence zones* could be different for homogeneous and HD traffic conditions (and explored three different lengths of the rectangular influence area for these traffic conditions). However, *influence zones* can potentially depend on various other factors such as the type of the subject vehicle, driver's characteristics and human factors, subject vehicle speed, and traffic stream characteristics such as average speed and density. One approach to recognise such heterogeneity and dynamic nature of *influence zones* is to use a latent class framework, where different sizes/shapes of *influence zones* are probabilistically associated with the subject vehicle and driver characteristics, macroscopic traffic stream variables, and microscopic (time-dependent) variables such as subject vehicle speed. The driving decisions can be modelled conditional on the (latent) *influence zones*. To do so, however, empirical data would be needed from a variety of traffic conditions (representing different ranges of traffic speeds and densities), vehicle types, etc. Also, one can explore an elliptical shape or a conical shape for the influence zone.

8.4.3 Comparison of Driver Behaviour in Homogeneous and HD Traffic Streams

The insights we obtained on the differences in car driver behaviour between HD and *homogeneous traffic* streams may be not only due to different traffic conditions in the two types of streams. The differences may also be due to other factors such as driving population characteristics, vehicle characteristics, geometry and types of roadway facilities, congestion levels, weather, and timing of data collection. The panel data models proposed in Chapter 4 help control for the confounding effects of unobserved factors (e.g., vehicle- and driver-specific factors) that vary across different vehicles in each of the two datasets. However, the models do not help control for factors that are different between the two datasets but are not different across the vehicles within each dataset. These include, for example, congestion levels and traffic composition in the two traffic streams, roadway geometry, and type of facilities. To control for the effects of such factors when comparing driving behaviour between HD and *homogeneous traffic* conditions, it is important to analyse a greater variety of trajectory datasets from a larger number of locations representing variation in such factors in both HD traffic and *homogeneous traffic* conditions. Until then, the findings on differences in driving behaviour ought to be used with caution. Another possibility is to explore driving simulator experiments to analyse differences in driving behaviour between HD traffic and *homogeneous traffic* conditions while controlling for other possible confounding factors.

8.4.4 Disentangling the Variability due to Perception Errors in Traffic Environment Variables from the Heterogeneity in Response to these Variables

Chapter 5 proposed a discrete choice modelling framework that accommodates perception errors in choice environment variables that do not vary across choice alternatives. However, note that, due to the mathematical equivalence of the proposed modelling framework with the mixed logit model, random parameters could be confounded with the errors in the attributes.

Such stochasticity in variables is different from unobserved heterogeneity in drivers' responses to those traffic environment variables. Specifying only one of these as random could potentially lead to biased parameter estimates. Simultaneous identification of these two from separate sources of variability is important in understanding driver behaviour in complex traffic environments. Díaz et al. (2015) pointed out that it is very difficult to isolate the unobserved heterogeneity from stochasticity in attributes. Hence, it would be useful to explore ways to separate the two sources of variability.

8.4.5 2D Movement Models for Cars and Autorickshaws

While modelling 2D movement, this dissertation only focused on motorised two-wheelers. The modelling framework formulated in Chapter 6 can be employed to model the 2D movement of cars and autorickshaws in HD traffic conditions. Also, the formulated framework did not consider driver's perception error in traffic environment variables. The incorporation of perception error would help increase the realism of the 2D driver behaviour model.

8.4.6 Consideration of Human Factors while Modelling Driver Behaviour

This dissertation only focuses on two human factors – MVA and drivers' perception errors – while modelling driver behaviour. Other human factors, such as aggressiveness and carefulness in driving, driving behaviours based on anticipating future actions of other vehicles, etc., are not adequately addressed in this dissertation. As importantly, this dissertation reduces driving behaviour to the decisions drivers make in the next time step based on the current traffic environment. By doing so, the time dynamics of driving over several time steps, such as the influence of one's anticipated future actions on current actions, are not considered.

8.4.7 Traffic Simulator for HD traffic streams

The developed traffic simulator in this dissertation has several limitations for simulating HD traffic streams. The developed models only consider longitudinal movements for cars and 2D movements for motorised two-wheelers. Hence, the current traffic simulator only simulated the longitudinal movement of cars and the 2D movements of motorised two-wheelers. Furthermore, the traffic simulator did not simulate other vehicle types such as auto-rickshaws, buses, heavy vehicles, and light commercial vehicles. These vehicle types constitute a good amount of vehicle composition in HD traffic streams, and hence important to simulate them as well. Moreover, we observed collisions of vehicles a few times while simulating HD traffic streams using the developed traffic simulator. Hence, some heuristics are used to avoid collisions. These issues must be addressed in future work.

8.4.8 Consideration of Drivers' Useful Visual Field

As per Mackworth (1965), *"the useful visual field can be defined as the area around the fixation point (the point in space on which the eyes are focused) inside which information can be perceived."* If the quantity of information to be processed increases, the useful visual field decreases (Rogé et al., 2004). For example, the useful visual field decreases if the number of vehicles within the visual field increases. Moreover, it decreases with the increase in speed. This indicates that the useful visual field will have a strong bearing on the behaviour of drivers because it directly influences the stimulus perception by drivers. Therefore, future endeavours in building driver behaviour models (specifically MVA-based models) shall consider this important factor.

8.4.9 Data Needs

We did not find studies focussing on collecting data specifically to capture MVA behaviour, probably because most of the time MVA based models are simple extensions of single lead vehicle-based models. As revealed in Chapter 2, MVA is not a straightforward process. Before incorporating it, one needs to answer how many vehicles a driver considers? What are the driver's focus areas? How to quantify a driver's useful visual field? How does the useful visual field vary for a particular driver? How does MVA change based on the type of subject vehicle (bus or car or motorised two-wheelers) and on vehicle classes in front? To answer all these questions, it is important to collect detailed driver behaviour data. For instance, researchers shall focus on collecting data up with a trap length of 100 m to 500 m road stretch so that kinematics of at least a few additional vehicles ahead can be captured. Notably, driving simulators offer various advantages, specifically the collection of detailed driver-level data. Various scenarios with a varying number of vehicles ahead, different vehicle classes, varied placement of vehicles in the driving scene, etc., can be created and tested. Next, eye trackers can be utilised to identify driver focus points within the visual field. Combining driving simulators and eye-tracking data can assist in a better understanding and modelling of the MVA behaviour.

A. APPENDIX TO CHAPTER 3

Assuming $h = \dfrac{m_{qi} - \alpha_i^T z_{qi}}{\sigma_{\eta i}}$, the conditional probability in Eq. (3.6) can be expressed as:

$$\begin{aligned}
P\left(m_{qi} | v_{qi} > -\beta_i^T x_{qi}\right) = P\left(\eta_{qi} = h \,|\, v_{qi} > -\beta_i^T x_{qi}\right) \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} P\left(v_{qi} > -\beta_i^T x_{qi}, \eta_{qi} = h\right) \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} P\left(v_{qi} > -\beta_i^T x_{qi} \,|\, \eta_{qi} = h\right) P\left(\eta_{qi} = h\right) \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} \left[1 - P\left(v_{qi} < -\beta_i^T x_{qi} \,|\, \eta_{qi} = h\right)\right] P\left(\eta_{qi} = h\right) \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} \left[P\left(\eta_{qi} = h\right) - P\left(v_{qi} < -\beta_i^T x_{qi} \,|\, \eta_{qi} = h\right) P\left(\eta_{qi} = h\right)\right] \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} \left[P\left(\eta_{qi} = h\right) - P\left(v_{qi} < -\beta_i^T x_{qi}, \eta_{qi} = h\right)\right]
\end{aligned} \tag{A.1}$$

The derivation for the expression given in Eq. (3.15) is as follows:

$$\begin{aligned}
&P\left(\eta_{qi} = h \,|\, v_{qi} > -\beta_i^T x_{qi}\right) \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} \left[\frac{1}{\sigma_{\eta i}} f_{\eta i}\left(h\right) - \frac{1}{\sigma_{\eta i}} \frac{\partial}{\partial t} F_{v_{qi}, \eta_{qi}}\left(-\beta_i^T x_{qi}, h\right)\right] \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} \left[\frac{1}{\sigma_{\eta i}} f_{\eta i}\left(h\right) - \frac{1}{\sigma_{\eta i}} f_{\eta i}(h) \frac{\partial}{\partial u_2^i} C_\theta\left(u_{q1}^i = F_{v_{qi}}\left(-\beta_i^T x_{qi}\right), u_{q2}^i = F_{\eta_{qi}}(h)\right)\right] \\
&= \left[P\left(v_{qi} > -\beta_i^T x_{qi}\right)\right]^{-1} \left[\frac{1}{\sigma_{\eta i}} f_{\eta i}\left(h\right) - \frac{1}{\sigma_{\eta i}} f_{\eta i}(h) \frac{\partial C_\theta\left(u_{q1}^i, u_{q2}^i\right)}{\partial u_2^i}\right]
\end{aligned} \tag{A.2}$$

The derivation for marginal CDF function of $F_{v_{qi}}$ is as follows:

$$\begin{aligned}
F_{v_{qi}}\left(-\beta_i^T x_{qi}\right) &= \Pr\left(v_{qi} < -\beta_i^T x_{qi}\right) \\
&= 1 - \Pr\left(v_{qi} > -\beta_i^T x_{qi}\right)
\end{aligned} \tag{A.3}$$

As defined earlier in the chapter, $v_{qi} = \varepsilon_{qi} - \left\{\max_{j=a,d,s,j\neq i} u_{qj}^*\right\}$. Therefore, we can write:

$$\Pr\left(v_{qi} > -\beta_i^T x_{qi}\right) = \Pr\left(\varepsilon_{qi} - \left\{\max_{j=a,d,s,j\neq i} u_{qj}^*\right\} > -\beta_i^T x_{qi}\right) \tag{A.4}$$

Assume that the error terms ε_{qj} are IID Gumbel distributed with location parameter 0 and scale parameter σ_G (i.e., $\varepsilon_{qj} \sim \text{Gumbel}(0, \sigma_G)$). Given this assumption, $\left\{\max_{j=a,d,s,j\neq i} u_{qj}^*\right\}$, which represents maximum of IID Gumbel random variables, is also Gumbel distributed with the same scale parameter σ_G. Specifically, following Ben-Akiva and Lerman (1985), one can write $\left\{\max\limits_{j=a,d,s,j\neq i} u^*_{qj}\right\} = \frac{1}{\sigma_G}\ln\sum\limits_{j\neq i}\exp\left(\sigma_G\beta_j^T x_{qj}\right) + \varepsilon_j^*$, where $\varepsilon_j^* \sim \text{Gumbel}(0, \sigma_G)$. Substituting this expression into Eq. (A.4), one can write the following:

$$
\begin{aligned}
\Pr\left(v_{qi} > -\beta_i^T x_{qi}\right) &= \Pr\left(\varepsilon_{qi} - \left\{\max_{j=a,d,s,j\neq i} u^*_{qj}\right\} > -\beta_i^T x_{qi}\right) \\
&= \Pr\left(\varepsilon_{qi} - \frac{1}{\sigma_G}\ln\sum_{j\neq i}\exp\left(\sigma_G\beta_j^T x_{qj}\right) - \varepsilon_j^* > -\beta_i^T x_{qi}\right) \\
&= \Pr\left(\varepsilon_j^* - \varepsilon_{qi} < \beta_i^T x_{qi} - \frac{1}{\sigma_G}\ln\sum_{j\neq i}\exp\left(\sigma_G\beta_j^T x_{qj}\right)\right)
\end{aligned}
\tag{A.5}
$$

Since ε_{qi} and ε_j^* are identically Gumbel distributed, $\varepsilon_{qi} - \varepsilon_j^*$ follows a logistic distribution (Ben-Akiva and Lerman, 1985). Therefore, the above expression can be simplified as:

$$
\Pr\left(v_{qi} > -\beta_i^T x_{qi}\right) = \frac{\exp\left(\sigma_G\beta_i^T x_{qi}\right)}{\exp\left(\sigma_G\beta_i^T x_{qi}\right) + \exp\left(\ln\sum\limits_{j\neq i}\exp\left(\sigma_G\beta_j^T x_{qj}\right)\right)} \tag{A.6}
$$

Since we normalize $\sigma_G = 1$ for identification, one can write $F_{v_{qi}}\left(-\beta_i^T x_{qi}\right)$ or u^i_{q1} as below:

$$
u^i_{q1} = 1 - \frac{\exp\left(\beta_i^T x_{qi}\right)}{\exp\left(\beta_i^T x_{qi}\right) + \exp\left(\ln\sum\limits_{j\neq i}\exp\left(\beta_j^T x_{qj}\right)\right)} \tag{A.7}
$$

B. APPENDIX TO CHAPTER 4

Table B.1 Estimation results of the joint models on homogeneous traffic dataset for influence zones of lengths 45 m and 60 m

Explanatory variables	Influence zone length ahead of SV = 45 m				Influence zone length ahead of SV = 60 m			
	Discrete choice#		Continuous choice*		Discrete choice#		Continuous choice*	
	Accn	**Dccn**	**Accn**	**Dccn**	**Accn**	**Dccn**	**Accn**	**Dccn**
Constant	-0.156 (-1.25)	-0.120 (-0.79)	0.906 (16.47)	1.122 (14.42)	-0.114 (-0.91)	-0.065 (-0.42)	0.904 (16.57)	1.136 (14.36)
Subject vehicle's speed in longitudinal direction (m/s)	0.024 (1.60)	0.075 (5.65)	--	0.011 (1.84)	0.020 (1.35)	0.073 (5.58)	--	0.011 (1.74)
Space gap in longitudinal direction w.r.t. MF1 (m)	0.020 (3.54)	-0.021 (-3.52)	0.005 (2.57)	-0.003 (-1.15)	0.019 (3.35)	-0.023 (-3.88)	0.005 (2.58)	-0.005 (-1.76)
Relative speed in longitudinal direction w.r.t. MF1 (m/s)	0.411 (13.30)	-0.343 (-11.43)	0.065 (5.49)	-0.057 (-4.54)	0.419 (13.26)	-0.333 (-10.87)	0.061 (5.03)	-0.055 (-4.34)
Subject vehicle has 2 or more lead vehicles in MF compartment**	--	--	--	0.145 (1.82)	--	--	--	0.061 (0.90)
Space gap in longitudinal direction w.r.t. MF2 (m)	--	--	--	-0.006 (-2.51)	--	--	--	-0.002 (-1.14)
Relative speed in longitudinal direction w.r.t. MF2 (m/s)	0.070 (2.79)	-0.200 (-7.67)	0.014 (1.55)	-0.048 (-4.90)	0.040 (1.77)	-0.186 (-7.76)	0.018 (2.30)	-0.042 (-4.98)
Subject vehicle has 1 or more lead vehicles in LF compartment***	--	--	--	--	--	--	--	--
Space gap in longitudinal direction w.r.t. LF1 (m)	--	--	--	--	--	--	--	--
Lateral gap between MF1 and LF1 (m)	0.018 (0.83)	--	--	-0.010 (-1.08)	0.020 (0.94)	--	--	-0.009 (-0.88)
Relative speed in longitudinal direction w.r.t. LF1 (m/s)	--	-0.007 (-1.08)	--	--	--	-0.007 (-1.09)	--	--
Subject vehicle has 1 or more lead vehicles in RF compartment***	--	--	--	--	--	--	--	--
Space gap in longitudinal direction w.r.t. RF1 (m)	--	--	--	--	--	--	--	--
Lateral gap between MF1 and RF1 (m)	--	--	--	--	--	--	--	--
Relative speed in longitudinal direction w.r.t. RF1 (m/s)	0.015 (1.34)	--	--	-0.006 (-1.21)	0.014 (1.28)	--	--	-0.005 (-1.02)
Subject vehicle has 1 or more side vehicle in LS compartment***	--	--	--	--	--	--	--	--
Space gap in lateral direction w.r.t. LS1 (m)	--	--	--	--	--	--	--	--
Relative speed in longitudinal direction w.r.t. LS1 (m/s)	--	--	--	--	--	--	--	--
Subject vehicle has 1 or more side vehicle in RS compartment***	--	--	--	--	--	--	--	--
Space gap in lateral direction w.r.t. RS1 (m)	--	--	--	--	--	--	--	--
Relative speed in longitudinal direction w.r.t. RS1 (m/s)	--	--	--	--	--	--	--	--
Space gap between left edge of the SV and left edge of the road (m)	--	0.011 (2.06)	--	--	--	0.010 (1.88)	--	--
Scale parameter of regression equations $(\sigma_{\eta i})$			0.625 (43.65)	0.737 (53.38)			0.624 (43.88)	0.739 (54.29)
Scale parameters for panel effects $(\sigma_{\psi i}$ and $\sigma_{\xi i})$								
Acceleration	0.206 (3.01)		--		0.208 (3.05)		--	
Deceleration	0.182 (2.82)		--		0.189 (3.02)		--	
Maintain same speed	0.137 (1.36)		NA		0.133 (1.29)		NA	
Scale parameters of driver-level common error terms $(\sigma_{\omega i})$								
Acceleration	0.065 (2.76)				0.065 (2.79)			
Deceleration	--				--			
Copula dependency parameters (θ_i)								
Acceleration	-2.871 (-6.29)				-2.853 (-6.25)			
Deceleration	-4.377 (-10.85)				-4.383 (-10.99)			
Goodness of fit measures								
Number of parameters	35				35			
Log likelihood	-14605.60				-14622.28			
AIC value	29281.20				29314.56			
BIC value	29528.80				29562.16			
LLR value w.r.t. to independent model	351.24				317.87			
Critical chi-square value at 95% CI	7.81				7.81			
Adjusted rho-square	0.102				0.101			
Number of cases	8728				8728			
Number of vehicles	522				522			

Notes: Accn – Acceleration, Dccn – Deceleration, # Maintain same speed is base, *Dependent variable = absolute value of acceleration/deceleration at t s (m/s2), ** One lead vehicle is base, *** No vehicle is base, -- the corresponding parameter was dropped from the specification as it was found to be statistically insignificant, NA- Not applicable

Table B.2 Estimation results of the joint models on HD traffic dataset for influence zones of lengths 30 m and 45 m

Explanatory variables	Influence zone length ahead of SV = 30 m				Influence zone length ahead of SV = 45 m			
	Discrete choice#		Continuous choice*		Discrete choice#		Continuous choice*	
	Accn	Dccn	Accn	Dccn	Accn	Dccn	Accn	Dccn
Constant	2.246 (5.88)	-0.850 (-2.29)	1.257 (14.10)	-0.055 (-0.57)	2.018 (5.34)	-0.882 (-2.32)	1.182 (12.95)	-0.068 (-0.70)
Subject vehicle's speed in longitudinal direction (m/s)	-0.154 (-4.71)	0.211 (7.26)	-0.029 (-3.55)	0.078 (12.04)	-0.143 (-4.39)	0.212 (7.38)	-0.025 (-3.00)	0.079 (12.04)
Space gap in longitudinal direction w.r.t. MF1 (m)	0.029 (4.12)	-0.014 (-2.09)	0.004 (2.22)	-0.004 (-2.58)	0.025 (3.66)	-0.016 (-2.36)	0.004 (2.04)	-0.004 (-2.63)
Relative speed in longitudinal direction w.r.t. MF1 (m/s)	0.090 (4.66)	-0.073 (-3.58)	0.026 (5.46)	-0.004 (-0.77)	0.083 (4.38)	-0.077 (-3.89)	0.025 (5.31)	-0.004 (-0.79)
Subject vehicle has 2 or more lead vehicles in MF compartment**	--	0.176 (1.99)	--	--	--	0.220 (2.81)	--	--
Space gap in longitudinal direction w.r.t. MF2 (m)	0.006 (1.62)	--	--	--	0.010 (4.10)	--	0.001 (1.09)	--
Relative speed in longitudinal direction w.r.t. MF2 (m/s)	0.070 (2.41)	-0.036 (-1.33)	0.017 (2.83)	--	0.080 (5.69)	--	0.016 (3.11)	--
Subject vehicle has 1 or more lead vehicles in LF compartment***	--	0.306 (2.72)	--	--	--	0.252 (1.92)	--	--
Space gap in longitudinal direction w.r.t. LF1 (m)	--	-0.016 (-3.46)	0.003 (1.56)	--	--	-0.011 (-2.94)	0.002 (1.54)	--
Lateral gap between MF1 and LF1 (m)	0.106 (5.12)	--	--	--	0.108 (5.13)	--	--	--
Relative speed in longitudinal direction w.r.t. LF1 (m/s)	0.058 (3.20)	-0.022 (-1.26)	--	--	0.055 (3.15)	-0.021 (-1.24)	--	--
Subject vehicle has 1 or more lead vehicles in RF compartment***	--	0.204 (1.89)	-0.100 (-2.96)	0.088 (3.61)	--	0.140 (1.29)	-0.097 (-2.95)	0.093 (3.71)
Space gap in longitudinal direction w.r.t. RF1 (m)	--	-0.012 (-2.33)	0.002 (1.27)	--	--	-0.008 (-2.13)	0.004 (3.16)	--
Lateral gap between MF1 and RF1 (m)	--	--	--	--	--	--	--	--
Relative speed in longitudinal direction w.r.t. RF1 (m/s)	--	-0.064 (-4.78)	0.021 (4.15)	--	--	-0.056 (-4.75)	0.020 (4.19)	--
Subject vehicle has 1 or more side vehicle in LS compartment***	--	--	--	--	--	--	--	--
Space gap in lateral direction w.r.t. LS1 (m)	0.050 (2.45)	--	--	--	0.055 (2.66)	--	--	--
Relative speed in longitudinal direction w.r.t. LS1 (m/s)	--	-0.035 (-3.04)	--	--	--	-0.035 (-2.95)	--	--
Subject vehicle has 1 or more side vehicle in RS compartment***	-0.172 (-2.23)	--	--	--	-0.169 (-2.20)	--	--	--
Space gap in lateral direction w.r.t. RS1 (m)	--	--	--	--	--	--	--	--
Relative speed in longitudinal direction w.r.t. RS1 (m/s)	0.050 (2.63)	--	--	-0.019 (-3.27)	0.048 (2.50)	--	--	-0.018 (-3.17)
Space gap between left edge of the SV and left edge of the road (m)	--	-0.102 (-3.95)	--	--	--	-0.104 (-3.89)	--	--
Scale parameter of regression equations $(\sigma_{\eta i})$			0.608 (50.10)	0.579 (55.97)			0.606 (51.63)	0.578 (56.62)
Scale parameters for panel effects $(\sigma_{\psi i}$ and $\sigma_{\xi i})$								
Acceleration	0.543 (8.00)		--		0.552 (8.18)		--	
Deceleration	0.397 (4.59)		--		0.388 (4.41)		--	
Maintain same speed	--		NA		--		NA	
Scale parameters of driver-level common error terms $(\sigma_{\omega i})$								
Acceleration	0.143 (7.45)				0.145 (7.73)			
Deceleration	0.187 (12.39)				0.190 (12.58)			
Copula dependency parameters (θ_i)								
Acceleration	-5.205 (-11.19)				-5.267 (-11.43)			
Deceleration	-5.240 (-12.52)				-5.179 (-12.36)			
Goodness of fit measures								
Number of parameters	48				48			
Log likelihood	-10802.09				-10789.16			
AIC value	21700.18				21674.32			
BIC value	22028.56				22002.70			
LLR value w.r.t. to independent model	810.18				836.04			
Critical chi-square value at 95% CI	5.99				5.99			
Adjusted rho-square	0.140				0.141			
Number of cases	6914				6914			
Number of vehicles	760				760			

Accn – Acceleration, Dccn – Deceleration, # Maintain same speed is base, *Dependent variable = absolute value of acceleration/deceleration at t s (m/s2), ** One lead vehicle is base, *** No vehicle is base, -- the corresponding parameter was dropped from the specification as it was found to be statistically insignificant, NA- Not applicable

Table B.3 Comparison of goodness of fit measures for the models in Chapters 4 and 3

Goodness of fit measures	**Homogeneous traffic conditions**		**Heterogeneous disorderly traffic conditions**	
	Model proposed in Chapter 4	**Model proposed in Chapter 3**	**Model proposed in Chapter 4**	**Model proposed in Chapter 3**
Number of parameters	37	34	50	52
Log likelihood	-14554.46	-14574.88	-10768.14	-10931.94
AIC value	29182.93	29217.77	21636.28	21967.88
BIC value	29444.68	29458.29	21978.35	22323.63
LLR value w.r.t. to independent model	453.51	412.66	878.08	550.48
Critical chi-square value at 95% CI	11.07	5.99	9.49	12.59
Adjusted rho-square	0.11	0.10	0.14	0.13
Number of vehicles	522	522	760	760
Number of cases	8728	8728	6914	6914

C. APPENDIX TO CHAPTER 5

C.1 Estimation of the Proposed Mixed Logit Models with Multiplicative Errors

The parameters of the proposed ML-ME models can be estimated using the MSL estimation routine. To do so, building on Eq. (5.3) for the likelihood expression for an individual driver q 's manoeuvring choice, the likelihood expression for a sample of independent observations $(q = 1, 2, ..., Q)$ may be written as:

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

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

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

where, τ_q is a vector of all perception error terms $\tau_{qk}\,(k = 1, 2, ..., K)$ 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:

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

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

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

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, which is given in the expression below, is maximized to obtain the parameters (β,θ):

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

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. However, with 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. (B.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):

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

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:

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

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 D.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. Recall that the power lognormal distribution has a power parameter in addition to the variable coefficients and the scale parameter. The Weibull and the Fréchet distribution have an additional shape parameter. The gradients with respect to those parameters are not presented here, since those parameters were fixed for ease of estimation. Gauss mathematical modelling software was used to code the simulated likelihood functions and their gradients for parameter estimation.

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

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

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

C.2 Variance-Covariance Matrix for Additive Error Specification

Consider the additive error specification on traffic environment variables as in Equations (5.7) and (5.8) .

$$
\begin{aligned}
Var(\varepsilon_{qa}) &= Var(\beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa}) \\
&= E[(\beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa})^2] - (E[(\beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa})])^2 \\
&= \beta_{a1}^2\sigma_{\eta 1}^2 + \beta_{a2}^2\sigma_{\eta 2}^2 + \beta_{a3}^2\sigma_{\eta 3}^2 + \sigma_{\xi}^2
\end{aligned}
$$

$$
\begin{aligned}
Cov(\varepsilon_{qa}, \varepsilon_{qd}) &= E[(\varepsilon_{qa} - E(\varepsilon_{qa}))(\varepsilon_{qd} - E(\varepsilon_{qd}))] \\
&= E[(\varepsilon_{qa})(\varepsilon_{qd})] \\
&= E[(\beta_{a1}\eta_{q1} + \beta_{a2}\eta_{q2} + \beta_{a3}\eta_{q3} + \xi_{qa})(\beta_{d1}\eta_{q1} + \beta_{d2}\eta_{q2} + \beta_{d3}\eta_{q3} + \xi_{qd})] \\
&= \beta_{a1}\beta_{d1}\sigma_{\eta_1}^2 + \beta_{a2}\beta_{d2}\sigma_{\eta_2}^2 + \beta_{a3}\beta_{d3}\sigma_{\eta_3}^2
\end{aligned}
$$

Hence, the variance-covariance matrix can be expressed as follows:

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

As alternative 3 is the base alternative, the variance-covariance matrix of error differences can be calculated as $\Omega_{\Delta} = M\Omega M'$, where $M = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$.

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

C.3 Variance-covariance Matrix for Multiplicative Error Specification

Consider the multiplicative error specification on traffic environment variables as in Equations (5.14) and (5.15).

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

One can write the utility specification as:

$$\begin{aligned} U_{qa} &= \beta_{a0} + \varsigma_{qa} \\ U_{qd} &= \beta_{d0} + \varsigma_{qd} \\ U_{qs} &= \varsigma_{qs} \end{aligned}$$

The elements of the variance-covariance are derived next.

$$\begin{aligned} Var[\varsigma_{qa}] &= Var[\beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa}] \\ &= E[(\beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa})^2] \\ &\quad -(E[(\beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa})])^2 \end{aligned}$$

Let $A = E[(\beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa})^2]$ and

$B = (E[(\beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa})])^2$.

Then, $Var(\varsigma_{qa}) = A - B$.

$$
\begin{aligned}
A &= E[(\beta_{a1}(x_{q1}\tau_{q1}) + \beta_{a2}(x_{q2}\tau_{q2}) + \beta_{a3}(x_{q3}\tau_{q3}) + \xi_{qa})^2] \\
&= E[(\beta_{a1}x_{q1}\tau_{q1})^2 + (\beta_{a2}x_{q2}\tau_{q2})^2 + (\beta_{a3}x_{q3}\tau_{q3})^2 + (\xi_{qa})^2 \\
&\quad +2(\beta_{a1}x_{q1}\tau_{q1}\beta_{a2}x_{q2}\tau_{q2} + \beta_{a1}x_{q1}\tau_{q1}\beta_{a3}x_{q3}\tau_{q3} + \beta_{a1}x_{q1}\tau_{q1}\xi_{qa} \\
&\quad +\beta_{a2}x_{q2}\tau_{q2}\beta_{a3}x_{q3}\tau_{q3} + \beta_{a2}x_{q2}\tau_{q2}\xi_{qa} \\
&\quad +\beta_{a3}x_{q3}\tau_{q3}\xi_{qa})] \\
&= \beta_{a1}x_{1q}E[(\tau_{q1})^2] + \beta_{a2}x_{q2}E[(\tau_{q2})^2] + \beta_{a3}x_{q3}E[(\tau_{q3})^2] + E[(\xi_{qa})^2] \\
&\quad +2(\beta_{a1}x_{q1}\beta_{a2}x_{q2}E[\tau_{q1}\tau_{q2}] + \beta_{a1}x_{q1}\beta_{a3}x_{q3}E[\tau_{q1}\tau_{q3}] + \beta_{a1}x_{q1}E[\tau_{q1}\xi_{qa}] \\
&\quad +\beta_{a2}x_{q2}\beta_{a3}x_{q3}E[\tau_{q2}\tau_{q3}] + \beta_{a2}x_{q2}E[\tau_{q2}\xi_{qa}] \\
&\quad +\beta_{a3}x_{q3}E[\tau_{q3}\xi_{qa}])
\end{aligned}
$$

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, as $\tau_{qk} = x^*_{qk} / x_{qk}$, the random term should only take positive value. We assumed that τ_{qk} is lognormally distributed with mean one. Hence, $E[\tau_{q1}]=1$, $E[\tau_{q2}]=1$ and $E[\tau_{q1}]=1$. In our case, the expected value of a lognormally distributed error component will be equal to one if $\mu_{\tau k} = -\frac{1}{2}\sigma^2_{\tau k} \quad \forall k = 1,2,3$. This leads to the following conditions on the variances of the perception error terms:

$$
\begin{aligned}
&E[(\tau_{q1})^2] = \exp(\sigma^2_{\tau 1}) \\
&E[(\tau_{q1})^2] = \exp(\sigma^2_{\tau 2}) \\
&E[(\tau_{q1})^2] = \exp(\sigma^2_{\tau 3})
\end{aligned}
$$

Since the kernel error terms ξ_{qj} $(j = a,d,s)$ are assumed to be IID Gumbel error terms with mean zero and standard deviation σ_ξ, $E[(\xi_{qa})^2] = E[(\xi_{qd})^2] = E[(\xi_{qs})^2] = \sigma^2_\xi$.

Further, since τ_{qk} are IID lognormal distributed with mean one, one can write,

$$E[\tau_{q1}\tau_{q2}] = E[\tau_{q1}]E[\tau_{q2}] = 1$$

$$E[\tau_{q1}\tau_{q2}] = E[\tau_{q1}\tau_{q3}] = E[\tau_{q2}\tau_{q3}] = 1$$

$$E[\tau_{q1}\xi_{qa}] = E[\tau_{q1}]E[\xi_{qa}] = 1\times 0 = 0$$

$$E[\tau_{q1}\xi_{qa}] = E[\tau_{q2}\xi_{qa}] = E[\tau_{q3}\xi_{qa}] = 0$$

Therefore, $$A=(\beta_{a1}x_{q1})^2\exp(\sigma^2_{\tau 1})+(\beta_{a2}x_{q2})^2\exp(\sigma^2_{\tau 2})+(\beta_{a3}x_{q3})^2\exp(\sigma^2_{\tau 3})+\sigma^2_{\xi}$$ $$+2(\beta_{a1}x_{q1}\beta_{a2}x_{q2}+\beta_{a1}x_{q1}\beta_{a3}x_{q3}+\beta_{a1}x_{q1}\beta_{a3}x_{q3})$$ and

$$B=(\beta_{a1}x_{q1})^2+(\beta_{a2}x_{q2})^2+(\beta_{a3}x_{q3})^2+2(\beta_{a1}x_{q1}\beta_{a2}x_{q2}+\beta_{a1}x_{q1}\beta_{a3}x_{q3}+\beta_{a1}x_{q1}\beta_{a3}x_{q3})$$

Now, one can write $Var(\varsigma_{qa})=A-B$ as

$$Var(\varsigma_{qa})=(\beta_{a1}x_{q1})^2[\exp(\sigma^2_{\tau 1})-1]+(\beta_{a2}x_{q2})^2[\exp(\sigma^2_{\tau 2})-1]+(\beta_{a3}x_{q3})^2[\exp(\sigma^2_{\tau 3})-1]+\sigma^2_{\xi}.$$

Next, one can derive the following expression for the expected value of ς_{qa},

$$\begin{aligned}E[\varsigma_{qa}]&=E[\beta_{a1}(x_{q1}\tau_{q1})+\beta_{a2}(x_{q2}\tau_{q2})+\beta_{a3}(x_{q3}\tau_{q3})+\xi_{qa}]\\&=\beta_{a1}x_{q1}+\beta_{a2}x_{q2}+\beta_{a3}x_{q3}\end{aligned}$$

Therefore, one can write,

$$\begin{aligned}&E[\varsigma_{qa}]=\beta_{a1}x_{q1}+\beta_{a2}x_{q2}+\beta_{a3}x_{q3}\\&E[\varsigma_{qd}]=\beta_{d1}x_{q1}+\beta_{d2}x_{q2}+\beta_{d3}x_{q3}\\&E[\varsigma_{qs}]=0\end{aligned}$$

Covariances between each pair of the above terms can be derived as:

$$\begin{aligned}&Cov(\varsigma_{qa},\varsigma_{qd})==\beta_{a1}\beta_{d1}\left(x_{q1}\right)^2\exp(\sigma^2_{\tau 1})+\beta_{a2}\beta_{d2}\left(x_{q2}\right)^2\exp(\sigma^2_{\tau 2})+\beta_{a3}\beta_{d3}\left(x_{q3}\right)^2\exp(\sigma^2_{\tau 3})\\&Cov(\varsigma_{qa},\varsigma_{qs})=0\\&Cov(\varsigma_{qd},\varsigma_{qs})=0\end{aligned}$$

Given all the above terms, the variance-covariance matrix of the random utility components can be written as below:

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

where,

$$\begin{aligned}&Var(\varsigma_{qa})=(\beta_{a1}x_{q1})^2[\exp(\sigma^2_{\tau 1})-1]+(\beta_{a2}x_{q2})^2[\exp(\sigma^2_{\tau 2})-1]+(\beta_{a3}x_{q3})^2[\exp(\sigma^2_{\tau 3})-1]+\sigma^2_{\xi}\\&Var(\varsigma_{qd})=(\beta_{d1}x_{q1})^2[\exp(\sigma^2_{\tau 1})-1]+(\beta_{d2}x_{q2})^2[\exp(\sigma^2_{\tau 2})-1]+(\beta_{d3}x_{q3})^2[\exp(\sigma^2_{\tau 3})-1]+\sigma^2_{\xi}\\&Var(\varsigma_{qs})=\sigma^2_{\xi}\end{aligned}$$

As alternative 3 is the base alternative, one can get the error differenced variance-covariance matrix Ω_Δ as $\Omega_\Delta = M\Omega M'$, where $M = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$.

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

C.4 Alternative Distributions Explored for Representing Drivers' Perception Errors

Table C.2 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.

Table C.3 Estimation results of MNL, ML-RC-PLN, and ML-CRC-PLN models*

Explanatory variables in the utility functions	**MNL model**		**ML-RC-PLN model**		**ML-CRC-PLN model**	
	Acceleration utility	**Deceleration utility**	**Acceleration utility**	**Deceleration utility**	**Acceleration utility**	**Deceleration utility**
Constant	1.824 (7.87)	-0.481 (-2.21)	1.805 (7.51)	-0.792 (-2.98)	1.912 (7.38)	-0.937 (-4.13)
Subject vehicle (SV) longitudinal speed (m/s)	-0.065 (-3.31)	0.177 (9.89)	-0.062 (-3.08)	0.227 (8.77)	-0.077 (-3.73)	0.175 (9.78)
Traffic environment variables with respect to MF1 (first vehicle in MF) at t-0.5 s						
Space gap in longitudinal direction (m)	0.010 (1.88)	-0.016 (-3.21)	0.011 (2.15)	-2.825 (-6.48)	0.010 (1.71)	-0.017 (-3.15)
Relative speed in longitudinal direction (m/s)	0.087 (4.71)	-0.143 (-7.78)	0.075 (4.01)	-0.788 (-4.87)	-1.346 (-4.53)	-0.991 (-3.72)
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)	-0.670 (-4.37)	--	-0.702 (-4.31)	--	-0.706 (-3.88)	--
Space gap in longitudinal direction (m)	0.020 (3.01)	--	0.020 (2.81)	--	0.020 (2.66)	--
Relative speed in longitudinal direction (m/s)	0.112 (4.55)	-0.041 (-1.73)	0.099 (3.82)	-0.078 (-2.66)	0.114 (4.21)	-0.044 (-1.72)
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)	--	0.393 (4.13)	--	0.431 (3.79)	--	--
Space gap in longitudinal direction (m)	--	-0.011 (-3.10)	--	-0.013 (-3.07)	--	-0.006 (-1.54)
Lateral gap between MF1 and LF1 (m)	0.049 (2.03)	-0.034 (-1.33)	0.050 (2.05)	-0.040 (-1.40)	0.056 (2.90)	--
Relative speed in longitudinal direction (m/s)	0.071 (6.98)	--	0.077 (6.83)	--	0.077 (6.76)	--
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)	--	--	--	--	--	0.241 (4.43)
Space gap in longitudinal direction (m)	--	--	--	--	--	--
Lateral gap between MF1 and RF1 (m)	--	--	--	--	--	--
Relative speed in longitudinal direction (m/s)	0.058 (4.96)	--	0.062 (4.86)	--	0.059 (4.48)	--
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)	-0.139 (-1.80)	--	-0.181 (-2.17)	--	-0.121 (-1.35)	--
Lateral space gap (m)	0.067 (3.03)	--	0.082 (3.39)	--	0.071 (2.74)	--
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)	--	--	--	--	-0.113 (-1.69)	-0.014 (.)
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)	--	-0.084 (-6.09)	--	-0.107 (-6.23)	--	--

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

Table C.4 Estimated scale parameters of random coefficients in ML-RC-PLN and ML-CRC-PLN models*

Explanatory variables in the utility functions	ML-RC-PLN model		ML-CRC-PLN model	
	Acceleration utility	**Deceleration utility**	**Acceleration utility**	**Deceleration utility**
Constant	#	#	#	#
Subject vehicle (SV) longitudinal speed (m/s)	#	#	#	#
Traffic environment variables with respect to MF1 (first vehicle in MF) at t-0.5 s				
Space gap in longitudinal direction (m)	--	3.687 (4.19)**	--	--
Relative speed in longitudinal direction (m/s)	--	2.036 (3.69)**	3.553 (2.85)****	1.710 (2.89)****
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)	#	#	#	#
Space gap in longitudinal direction (m)	--	--	--	--
Relative speed in longitudinal direction (m/s)	--	--	--	--
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)	#	#	#	#
Space gap in longitudinal direction (m)	--	--	--	--
Lateral gap between MF1 and LF1 (m)	--	--	--	--
Relative speed in longitudinal direction (m/s)	--	--	--	--
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)	--	--	--	--
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)	#	#	#	#
Lateral space gap (m)	--	--	--	--
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)	#	#	#	#

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 2.5 and estimated value of rho (t- statistic) is 1.000 (79.50). -- the parameter was dropped from the specifications as it was insignificant, # the corresponding parameter was not considered as random coefficient.

REFERENCES

Abou-Zeid, M., Kaysi, I., and Al-Naghi, H. (2011). Measuring aggressive driving behavior using a driving simulator: An exploratory study. 3rd International Conference on Road Safety and Simulation.

Ahmed, K. I., Ben-Akiva, M., Koutsopoulos, H. N., and Mishalani, R. (1996). Models of freeway lane changing and gap acceptance behavior. *Transportation and Traffic Theory*, *13*, 501-515.

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

Ahn, K., and Rakha, H. (2009). A field evaluation case study of the environmental and energy impacts of traffic calming. *Transportation Research Part D: Transport and Environment*, *14*(6), 411-424.

Ali, Y., Zheng, Z., Haque, M. M., and Wang, M. (2019). A game theory-based approach for modelling mandatory lane-changing behaviour in a connected environment. *Transportation research part C: emerging technologies*, *106*, 220-242.

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

Amrutsamanvar, R. (2020). Modeling lateral movement decisions of powered two wheelers in disordered heterogeneous traffic conditions. *Transportation Letters*, 1-20.

Amrutsamanvar, R., Muthurajan, B. R., and Vanajakshi, L. D. (2021). Extraction and analysis of microscopic traffic data in disordered heterogeneous traffic conditions. *Transportation Letters*, *13*(1), 1-20.

Asaithambi, G., Kanagaraj, V., and Toledo, T. (2016). Driving behaviors: Models and challenges for non-lane based mixed traffic. *Transportation in Developing Economies*, *2*(2).

Asaithambi, G., and Shravani, G. (2017). Overtaking behaviour of vehicles on undivided roads in non-lane based mixed traffic conditions. *Journal of Traffic and Transportation Engineering (English Edition)*, *4*(3), 252-261.

Azam, M. S., Bhaskar, A., and Haque, M. M. (2020). Driving behaviour modelling in the context of heterogeneous traffic and poor lane discipline conditions: the state of the art and beyond. *Transportmetrica A: Transport Science*, 1-68.

Bando, M., Hasebe, K., Nakayama, A., Shibata, A., and Sugiyama, Y. (1995). Dynamical model of traffic congestion and numerical simulation. *Physical Review E*, *51*(2), 1035-1042.

Bando, M., Hasebe, K., Nakanishi, K., and Nakayama, A. (1998). Analysis of optimal velocity model with explicit delay. *Physical Review E*, *58*(5), 5429.

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.

Ben-Akiva, M. E., and Lerman, S. R. (1985). Discrete choice analysis: theory and application to travel demand. MIT press.

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* (pp. 550-559). CRC Press.

Bexelius, S. (1968). An extended model for car-following. *Transportation Research*, *2*(1), 13-21.

Bhat, C. R., and Castelar, S. (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B: Methodological*, *36*(7), 593-616.

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

Bhat, C. R., and Eluru, N. (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B: Methodological*, *43*(7), 749-765.

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.

Bhat, C. R. (2018). New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B: Methodological*, *109*, 238-256.

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.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, *99*(suppl 3), 7280-7287.

Brackstone, M., and McDonald, M. (1999). Car-following: a historical review. *Transportation Research Part F: Traffic Psychology and Behaviour*, *2*(4), 181-196.

Budhkar, A. K., and Maurya, A. K. (2017). Multiple-leader vehicle-following behavior in heterogeneous weak lane discipline traffic. *Transportation in developing economies*, *3*(2), 1-14.

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

Calvert, S. C., Schakel, W. J., and van Lint, J. W. C. (2020). A generic multi-scale framework for microscopic traffic simulation part II–Anticipation Reliance as compensation mechanism for potential task overload. *Transportation Research Part B: Methodological*, *140*, 42-63.

Cameron, A. C., Li, T., Trivedi, P. K., and Zimmer, D. M. (2004). Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. *The Econometrics Journal*, *7*(2), 566-584.

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., Agrawal, S., and Vasishtha, K. (2004). Microscopic modeling of driver behavior in uninterrupted traffic flow. *Journal of Transportation Engineering 130*(4), 438-451.

Chakroborty, P., and Das, A. (2017). Principles of transportation engineering. PHI Learning Pvt. Ltd.

Chakroborty, P., Maurya, A. K., and Vikram, D. (2019). Understanding and Modelling Disorderly Traffic Streams. *Journal of the Indian Institute of Science, 99*(4), 1-13.

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.

Chandler, R. E., Herman, R., and Montroll, E. W. (1958). Traffic dynamics: studies in car following. *Operations Research*, *6*(2), 165-184.

Chandra, S., and Shukla, S. (2012). Overtaking behavior on divided highways under mixed traffic conditions. *Procedia-Social Behavioral Sciences*, *43*, 313-322.

Chen , J., Shi, Z., and Hu, Y. (2012). Stabilization analysis of a multiple look-ahead model with driver reaction delays. *23*(06), 1250048.

Chen , J., Liu, R., Ngoduy, D., and Shi, Z. (2016). A new multi-anticipative car-following model with consideration of the desired following distance. *Nonlinear Dynamics*, *85*(4), 2705-2717.

Chen, X.-Q., Xie, W.-J., Shi, J., and Shi, Q.-X. (2010). Perturbation and stability analysis of the multi-anticipative intelligent driver model. *International Journal of Modern Physics C*, *21*(05), 647-668.

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.

Daganzo, C. F. (1995). The cell transmission model, part II: network traffic. *Transportation Research Part B: Methodological*, *29*(2), 79-93.

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

Das, S., and Maurya, A. K. (2018a). Multivariate analysis of microscopic traffic variables using copulas in staggered car-following conditions. *Transportmetrica A: transport science*, *14*(10), 829-854.

Das, S., and Maurya, A. K. (2018b). Modelling of motorised two-wheelers: a review of the literature. *Transport Reviews*, *38*(2), 209-231.

Das, S., Budhkar, A., Maurya, A. K., and Maji, A. (2019). Multivariate Analysis on Dynamic Car-Following Data of Non-lane-Based Traffic Environments. *Transportation in Developing Economies*, *5*(2), 1-13.

Delpiano, R., Herrera, J., Laval, J., and Coeymans-Avaria, J. (2019). A two-dimensional car-following model for two-dimensional traffic flow problems.

Dey, P. P., Chandra, S., and Gangopadhyay, S. (2008). Simulation of mixed traffic flow on two-lane roads. *Journal of Transportation Engineering*, *134*(9), 361-369.

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.

Ding, N., Jiao, N., Zhu, S., and Liu, B. (2019). Structural equations modeling of real-time crash risk variation in car-following incorporating visual perceptual, vehicular, and roadway factors. *133*.

Durbin, J., and Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, *37*(3/4), 409-428.

Edie, L. C. (1961). Car-following and steady-state theory for noncongested traffic. *Operations Research*, *9*(1), 66-76.

Edie, L. C. (1963). Discussion of traffic stream measurements and definitions. *Proceedings of the Second International Symposium on the Theory of Traffic Flow*, 139-154.

Edwards, C., Creaser, J., Caird, J., Lamsdale, A., and Chisholm, S. (2003). Older and younger driver performance at complex intersections: Implications for using perception-response time and driving simulation.

Eissfeldt, N., and Wagner, P. (2003). Effects of anticipatory driving in a traffic flow model. *The European Physical Journal B-Condensed Matter and Complex Systems*, *33*(1), 121-129.

Embrechts, P., McNeil, A., and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. *Risk Management: Value at Risk and Beyond*, *1*, 176-223.

Farhi, N., Haj-Salem, H., and Lebacque, J. P. (2012). Multianticipative piecewise-linear car-following model. *Transportation Research Record: Journal of the Transportation Research Board*, *2315*(1), 100-109.

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.

Gartner, N. H., Messer, C. J., and Rathi, A. (2001). Traffic flow theory-A state-of-the-art report: revised monograph on traffic flow theory.

Gazis, D. C., Herman, R., and Potts, R. B. (1959). Car-following theory of steady-state traffic flow. *Operations Research*, *7*(4), 499-505.

Gazis, D. C., Herman, R., and Rothery, R. W. (1961). Nonlinear follow-the-leader models of traffic flow. *Operations Research*, *9*(4), 545-567.

Ge, H. X., Dai, S. Q., Dong, L. Y., and Xue, Y. (2004). Stabilization effect of traffic flow in an extended car-following model based on an intelligent transportation system application. *Physical Review E*, *70*(6), 066134.

Genest, C., and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, *12*(4), 347-368.

Gipps, P. G. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, *15*(2), 105-111.

Gipps, P. G. (1986). A model for the structure of lane-changing decisions. *Transportation Research Part B: Methodological*, *20*(5), 403-414.

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.

Gujarati, D. N., Porter, D. C., and Gunasekar, S. (2012). Basic econometrics. Tata mcgraw-hill education.

Hamdar, S. (2012). Driver Behavior Modeling. In A. Eskandarian (Ed.), *Handbook of Intelligent Vehicles* (pp. 537-558). Springer London.

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.

Hasebe, K., Nakayama, A., and Sugiyama, Y. (2003). Dynamical model of a cooperative driving system for freeway traffic. *Physical review E*, *68*(2), 026102.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271.

HCM. (2000). Highway Capacity Manual. The National Academies Press.

Hellerstein, D. (2005). Modeling Discrete Choice with Uncertain Data: An Augmented MNL Estimator. *American Journal of Agricultural Economics*, *87*(1), 77-84.

Helly, W. (1959). Simulation of bottlenecks in single-lane traffic flow. Proceedings of the Symposium on Traffic Flow Theory, Research Laboratories, General Motors, 207-238.

Herman, R., Montroll, E. W., Potts, R. B., and Rothery, R. W. (1959). Traffic dynamics: analysis of stability in car following. *Operations research*, *7*(1), 86-106.

Herman, R., and Potts, R. B. (1959). Single lane traffic theory and experiment. Proceedings of the Symposium on Traffic Flow Theory, Research Laboratories, General Motors, 120-146.

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.

Hidas, P. (2002). Modelling lane changing and merging in microscopic traffic simulation. *Transportation Research Part C: Emerging Technologies*, *10*(5-6), 351-371.

Hidas, P. (2005). Modelling vehicle interactions in microscopic simulation of merging and weaving. *Transportation Research Part C: Emerging Technologies*, *13*(1), 37-62.

Hoogendoorn, S., and Ossen, S. (2006). Empirical analysis of two-leader car-following behavior. *European Journal of Transport and Infrastructure Research*, *6*(3), 229-246.

Hoogendoorn, S., Ossen, S., and Schreuder, M. (2006). Empirics of multianticipative car-following behavior. *Transportation Research Record*, *1965*(1), 112-120.

Hoogendoorn, S., Hoogendoorn, R. G., 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.

Hu, Y., Ma, T., and Chen, J. (2014). An extended multi-anticipative delay model of traffic flow. *Communications in nonlinear science and numerical simulation*, *19*(9), 3128-3135.

Jin , S., Wang, D., Tao, P., and Li, P. (2010). Non-lane-based full velocity difference car following model. *Physica A: Statistical Mechanics and its Applications, 389*(21), 4654-4662.

Jin, Y., Xu, M., and Gao, Z. (2011). KdV and kink-antikink solitons in an extended car-following model. *Nonlinear Dynamics*, *6*(1), 011018.

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.

Kanagaraj, V., and Treiber, M. (2018). Self-driven particle model for mixed traffic and other disordered flows. *Physica A: Statistical Mechanics and its Applications*, *509*, 1-11.

Kao, C., and Schnell, J. F. (1987). Errors in variables in the multinomial response model. *Economics Letters*, *25*(3), 249-254.

Kesting, A., Treiber, M., and Helbing, D. (2007). General lane-changing model MOBIL for car-following models. *Transportation Research Record: Journal of the Transportation Research Board*(1999), 86-94.

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.

Krauss, S., Wagner, P., and Gawron, C. (1996). Continuous limit of the Nagel-Schreckenberg model. *Physical Review E*, *54*(4), 3707.

Krauss, S. (1998). Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics. *PhD thesis*.

Lajunen, T., and Parker, D. (2001). Are aggressive people aggressive drivers? A study of the relationship between self-reported general aggressiveness, driver anger and aggressive driving. *Accident Analysis & Prevention*, *33*(2), 243-255.

Lee, T.-C. (2007). An agent-based model to simulate motorcycle behaviour in mixed traffic flow. *PhD thesis*, Imperial College London (University of London).

Lee, T.-C., Polak, J. W., and Bell, M. G. (2009). New approach to modeling mixed traffic containing motorcycles in urban areas. *Transportation Research Record*, *2140*(1), 195-205.

Lenz, H., Wagner, C., and Sollacher, R. (1999). Multi-anticipative car-following model. *The European Physical Journal B-Condensed Matter and Complex Systems*, *7*(2), 331-335.

Li , L., Chen , X., and Li , Z. (2013). Asymmetric stochastic Tau Theory in car-following. *Transportation research part F: traffic psychology and behaviour*, *18*, 21-33.

Li, Y., Sun, D., Liu, W., Zhang, M., Zhao, M., Liao, X., and Tang, L. (2011). Modeling and simulation for microscopic traffic flow based on multiple headway, velocity and acceleration difference. *Nonlinear Dynamics*, *66*, 15-28.

Li, Y., Zhang, L., Peeta, S., Pan, H., Zheng, T., Li, Y., and He, X. (2015). Non-lane-discipline-based car-following model considering the effects of two-sided lateral gaps. *Nonlinear Dynamics*, *80*(1), 227-238.

Li, Y., Zhang, L., Zhang, B., Zheng, T., Feng, H., and Li, Y. (2016). Non-lane-discipline-based car-following model considering the effect of visual angle. *Nonlinear Dynamics*, *85*(3), 1901-1912.

Li , Z., and Liu , Y. (2006). Analysis of stability and density waves of traffic flow model in an ITS environment. *The European Physical Journal B-Condensed Matter and Complex Systems*, *53*(3), 367-374.

Mackworth, N. H. (1965). Visual noise causes tunnel vision. *Psychonomic Science*, *3*(1), 67-68.

Mahapatra, G., Maurya, A. K., and Chakroborty, P. (2018). Parametric study of microscopic two-dimensional traffic flow models: A literature review. *Canadian Journal of Civil Engineering*, *45*(11), 909-921.

Mallikarjuna, C., and Rao, K. R. (2011). Heterogeneous traffic flow modelling: a complete methodology. *Transportmetrica*, *7*(5), 321-345.

Manski, C. F. (1977). The structure of random utility models. *Theory and Decision*, *8*(3), 229-254.

Mathew, T. V., Munigety, C. R., and Bajpai, A. (2013). Strip-based approach for the simulation of mixed traffic conditions. *Journal of Computing in Civil Engineering*, *29*(5).

Maurya, A. (2007a). Development of a comprehensive microscopic model for simulation of large uninterrupted traffic streams without lane discipline. *Indian Institute of Technology, Kanpur*.

Maurya, A. (2007b). Development of a comprehensive microscopic model for simulation of large uninterrupted traffic streams without lane discipline. *PhD thesis*, Indian Institute of Technology, Kanpur

May, A., and Keller, H. E. (1967). Non-integer car-following models. *Highway Research Record, 199*(1), 19-32.

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.

Montanino, M., and Punzo, V. (2013). Reconstructed NGSIM I80-1. http://www.multitude-project.eu/exchange/101.html

Montanino, M., and Punzo, V. (2015). Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns. *Transportation Research Part B: Methodological*, *80*, 82-106.

Moridpour, S., Sarvi, M., and Rose, G. (2010). Lane changing models: a critical review. *Transportation letters*, *2*(3), 157-173.

Munigety, C. R., and Mathew, T. V. (2016). Towards behavioral modeling of drivers in mixed traffic conditions. *Transportation in Developing Economies*, *2*(1), 6.

Newell, G. F. (1961). Nonlinear effects in the dynamics of car following. *Operations research*, *9*(2), 209-229.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, *27*(4), 289-303.

Ni, D. (2007). Determining Traffic-Flow Characteristics by Definition for Application in ITS. *IEEE Transactions on Intelligent Transportation Systems*, *8*(2), 181-187.

Ni, D. (2015). Traffic flow theory: Characteristics, experimental methods, and numerical techniques. Butterworth-Heinemann.

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.

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.

Ossen, S. (2008). Longitudinal driving behavior: theory and empirics. *PhD thesis*.

Oviedo-Trespalacios, O., Haque, M. M., King, M., and Washington, S. (2017). Effects of road infrastructure and traffic complexity in speed adaptation behaviour of distracted drivers. *Accident Analysis & Prevention*, *101*, 67-77.

Ozaki, H. (1993). Reaction and anticipation in the car-following behavior. Proceedings of 12th International Symposium on Theory of Traffic Flow and Transportation, 349-366.

Panwai, S., and Dia, H. (2007). Neural agent car-following models. *IEEE Transactions on Intelligent Transportation Systems*, *8*(1), 60-70.

Park, B. B., and Won, J. (2006). Microscopic simulation model calibration and validation handbook.

Paschalidis, E., Choudhury, C. F., and Hess, S. (2019). Combining driving simulator and physiological sensor data in a latent variable model to incorporate the effect of stress in car-following behaviour. *Analytic Methods in Accident Research*, *22*, 100089.

Peng, G., and Sun, D. (2010). A dynamical model of car-following with the consideration of the multiple information of preceding cars. *Physics Letters A*, *374*(15-16), 1694-1698.

Pipes, L. A. (1953). An operational analysis of traffic dynamics. *Journal of Applied Physics 24*(3), 274-281.

Punzo, V., Borzacchiello, M. T., and Ciuffo, B. (2011). On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C: Emerging Technologies*, *19*(6), 1243-1262.

Raju, N., Kumar, P., Jain, A., Arkatkar, S. S., and Joshi, G. (2018). Application of trajectory data for investigating vehicle behavior in mixed traffic environment. *Transportation Research Record*, *2672*(43), 122-133.

Raju, N., Arkatkar, S., and Joshi, G. (2021). Modeling following behavior of vehicles using trajectory data under mixed traffic conditions: an Indian viewpoint. *Transportation Letters*, *13*(9), 649-663.

Ravishankar, K., and Mathew, T. V. (2011). Vehicle-type dependent car-following model for heterogeneous traffic conditions. *Journal of transportation engineering*, *137*(11), 775-781.

Rogé, J., Pébayle, T., Lambilliotte, E., Spitzenstetter, F., Giselbrecht, D., and Muzet, A. (2004). Influence of age, speed and duration of monotonous driving task in traffic on the driver's useful visual field. *Vision Res*, *44*(23), 2737-2744.

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.

Saifuzzaman, M., Zheng, Z., Haque, M. M., and Washington, S. (2015). Revisiting the Task–Capability Interface model for incorporating human factors into car-following models. *Transportation research part B: methodological*, *82*, 1-19.

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.

Sarkar, N. C., Bhaskar, A., Zheng, Z., and Miska, M. P. (2020). Microscopic modelling of area-based heterogeneous traffic flow: Area selection and vehicle movement. *Transportation Research Part C: Emerging Technologies*, *111*, 373-396.

Sharma, A., Ali, Y., Saifuzzaman, M., Zheng, Z., and Haque, M. M. (2017a). Human factors in modelling mixed traffic of traditional, connected, and automated vehicles. International Conference on Applied Human Factors and Ergonomics, 262-273.

Sharma, A., Ali, Y., Saifuzzaman, M., Zheng, Z., and Haque, M. M. (2017b). Human factors in modelling mixed traffic of traditional, connected, and automated vehicles. International Conference on Applied Human Factors and Ergonomics, *591*, 262-273.

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.

Sharma, A., Zheng, Z., and Bhaskar, A. (2018). A pattern recognition algorithm for assessing trajectory completeness. *Transportation research part C: emerging technologies*, *96*, 432-457.

Sharma, A., Zheng, Z., Bhaskar, A., and Haque, M. M. (2019). Modelling car-following behaviour of connected vehicles with a focus on driver compliance. *Transportation research part B: methodological*, *126*, 256-279.

Sharma, A., Zheng, Z., Kim, J., Bhaskar, A., and Haque, M. M. (2020). Is an informed driver a better decision maker? A grouped random parameters with heterogeneity-in-means approach to investigate the impact of the connected environment on driving behaviour in safety-critical situations. *Analytic Methods in Accident Research*, *27*.

Sharma, A., Zheng, Z., Kim, J., Bhaskar, A., and Mazharul Haque, M. (2021). Assessing traffic disturbance, efficiency, and safety of the mixed traffic flow of connected vehicles and traditional vehicles by considering human factors. *Transportation Research Part C: Emerging Technologies*, *124*, 102934.

Shiomi, Y., Hanamori, T., Uno, N., and Shimamoto, H. (2012). Modeling traffic flow dominated by motorcycles based on discrete choice approach. *Proceedings of 1st LATSIS Conference*.

Siuhi, S. (2009). Parametric study of stimulus-response behavior incorporating vehicle heterogeneity in car-following models. *PhD thesis*, University of Nevada, Las Vegas.

Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, *9*(6), 449-460.

Spissu, E., Pinjari, A. R., Pendyala, R. M., and Bhat, C. R. (2009). A copula-based joint multinomial discrete–continuous model of vehicle type choice and miles of travel. *Transportation*, *36*(4), 403-422.

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

Steimetz, 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.

Subramanian, H. (1996). Estimation of car-following models. *PhD thesis*, Massachusetts Institute of Technology.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257-285.

Talebpour, A., Mahmassani, H. S., and Hamdar, S. H. (2015). Modeling lane-changing behavior in a connected environment: A game theory approach. *Transportation Research Part C: Emerging Technologies*, *59*, 216-232.

Tasca, L. (2000). A review of the literature on aggressive driving research. Ontario Advisory Group on Safe Driving Secretariat, Road User Safety Branch, Ontario Ministry of Transportation.

Thankappan, A., Tamut, Y., and Vanajakshi, L. D. (2010). Traffic stream modeling under heterogeneous traffic conditions (*Traffic and Transportation Studies 2010* (pp. 401-411).

Thankappan, A., and Vanajakshi, L. D. (2015). Development and application of a traffic stream model under heterogeneous traffic conditions. *Journal of The Institution of Engineers (India): Series A*, *96*(4), 267-275.

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

Toledo, T. (2007). Driving behaviour: models and challenges. *Transport Reviews*, *27*(1), 65-84.

Tong, Y. L. (1990). The multivariate normal distribution. Springer Science & Business Media.

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.

Transport Department, Government of India (1889). Rules of the road regulations.

Treiber, M., Hennecke, A., and Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations. *Physical review E*, *62*(2).

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., Kesting, A., and Helbing, D. (2007). Influence of reaction times and anticipation on stability of vehicular traffic flow. *Transportation Research Record*, *1999*(1), 23-29.

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

van Lint, H., Schakel, W., Tamminga, G., Knoppers, P., and Verbraeck, A. (2016). Getting the Human Factor into Traffic Flow Models: New Open-Source Design to Simulate Next Generation of Traffic Operations. *Transportation Research Record*, *2561*(1), 25-33.

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.

van Lint, J. W. C., and Calvert, S. C. (2018). A generic multi-level framework for microscopic traffic simulation-Theory and an example case in modelling driver distraction. *Transportation Research Part B: Methodological*, *117*, 63-86.

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.

Venkatachalam, T. A., and Gnanavelu, D. (2009). Concentration of heterogeneous road traffic (*Advanced Technologies* (pp. 503–530). Intech.

Vikram, D., Mittal, S., and Chakroborty, P. (2022). Stabilized finite element computations with a two-dimensional continuum model for disorderly traffic flow. *Computers & Fluids*, *232*, 105205.

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.

Wang, T., Gao, Z. Y., and Zhao, X. M. (2006). Multiple velocity difference model and its stability analysis. *Acta Physica Sinica*, *55*(2), 634.

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.

Wilson, R., Berg, P., Hooper, S., and Lunt, G. (2004). Many-neighbour interaction and non-locality in traffic models. *The European Physical Journal B-Condensed Matter and Complex Systems*, *39*(3), 397-408.

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.

Yang, Q., and Koutsopoulos, H. N. (1996). A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research Part C: Emerging Technologies*, *4*(3), 113-129.

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

Yu, L., Shi, Z., and Zhou, B. (2008). Kink–antikink density wave of an extended car-following model in a cooperative driving system. *Communications in Nonlinear Science Numerical Simulation*, *13*(10), 2167-2176.

Zhang, X. (2014). Empirical analysis of a generalized linear multianticipative car-following model in congested traffic conditions. *Journal of Transportation Engineering*, *140*(6), 04014018.

Zheng, Z. (2014). Recent developments and research needs in modeling lane changing. *Transportation research part B: methodological*, *60*, 16-32.

Zimmer, D. M., and Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, *24*(1), 63-76.