Improving transportation planning tools for the Global South: applying advanced modelling frameworks to address data issues and accommodate behavioural complexities

Submitted in accordance with the requirements for
the degree of Doctor of Philosophy
by

Khatun E Zannat

University of Leeds

September 2023

Intellectual property and publications

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The research conducted within the scope of this thesis has led to the development of six articles. Each article is listed below, including the corresponding chapter number in the thesis, and where relevant, a full reference (for the published article). A statement of the authors' contributions is also included for each paper.

1. The work in Chapter 2 of this thesis appeared in publication as follows:

Zannat, K., Choudhury, C. F., & Hess, S. (2022). Modelling departure time choice of car commuters in Dhaka, Bangladesh. Transportation research record, 2676(2), 247-262.

The authors confirm contribution to the paper as follows: study conception and design: KZ; data collection: KZ, CC; analysis: KZ; interpretation of results and model refinement: KZ, CC & SH; manuscript preparation: KZ. All authors reviewed the results and the responses and approved the final version of the manuscript.

2. The work in Chapter 3 of this thesis appeared in publication as follows:

Zannat, K. E., Choudhury, C. F., & Hess, S. (2024). Modelling time-of-travel preferences capturing correlations between departure times and activity durations. Transportation Research Part A: Policy and Practice, 184, 104081.

The authors confirm contribution to the paper as follows: study conception and design: KZ; data collection: KZ, CC; analysis: KZ; interpretation of results and model refinement: KZ, CC & SH; manuscript preparation: KZ. All authors reviewed the results and the responses and approved the final version of the manuscript.

3. The work in Chapter 6 of this thesis appeared in publication as follows:

Zannat, K. E., Laudan, J., Choudhury, C. F., & Hess, S. (2024). Developing an agent-based microsimulation for predicting the Bus Rapid Transit (BRT) demand in developing countries: A case study of Dhaka, Bangladesh. Transport Policy, 148, 92-106.

The authors confirm contribution to the paper as follows: study conception and design: KZ; data collection: KZ, CC; analysis and software: KZ, JL; interpretation of results and model refinement: KZ, JL, CC & SH; manuscript preparation: KZ. All authors reviewed the results and the responses and approved the final version of the manuscript.

4. The work in Chapter 4 of this thesis is the manuscript to be submitted shortly:

Zannat, K., Saxena, S., Choudhury, C. F., & Hess, S. Joint modelling of activity type, start time and duration using a bounded multiple discrete-continuous framework: A case study of Dhaka, Bangladesh

The authors confirm contribution to the paper as follows: study conception and design: KZ; data collection: KZ, CC; analysis and software: KZ, SS; interpretation of results and model refinement: KZ, SS, CC & SH; manuscript preparation: KZ. All authors reviewed the results and the responses and approved the final version of the manuscript.

5. The work in Chapter 5 of this thesis is the manuscript to be submitted shortly:

Zannat, K., Choudhury, C. F., & Hess, S, Carrasco, J-A. The role of satiation in activity participation: a comparative analysis of fractional logit (FMNL) model and multiple discrete-continuous extreme value (MDCEV) model.

The authors confirm contribution to the paper as follows: study conception and design: KZ; data collection: KZ, JAC; analysis: KZ; interpretation of results and model refinement: KZ, CC & SH; manuscript preparation: KZ. All authors reviewed the results and the responses and approved the final version of the manuscript.

6. The work in Chapter 7 of this thesis is under the second round of peer review in the IET Intelligent Transport Systems journal:

Zannat, K., Choudhury, C. F., Hess, S., & Watling, D Investigating the relative precision of GPS, GSM and CDR data for inferring spatiotemporal travel trajectories.

The authors confirm contribution to the paper as follows: study conception and design: KZ; data collection: KZ, CC; analysis: KZ; interpretation of results and model refinement: KZ, CC, SH, DW; manuscript preparation: KZ. All authors reviewed the results and the responses and approved the final version of the manuscript.

Abstract

Transport planning in the Global South faces unique challenges, particularly regarding data reliability and behavioural complexities. Nevertheless, since the transport planning literature is predominantly based on findings from high income countries, the planners and policymakers in the Global South often have to adopt transport models originally developed for the Global North. However, such models often overlook the behavioural dynamics and contextual factors specific to the Global South. This oversight can lead to uncertainties in results, and consequently, in the planning decisions. The primary barriers to developing effective transport planning tools in this context include the absence of reliable data sources, a lack of thorough investigation into behavioural complexity, and the absence of a framework for modelling choices in the Global South. This thesis aims to address some of the challenges related to data scarcity, behavioural complexity, and unique contextual factors in travel behaviour modelling by focusing on two distinct cities in the Global South: Dhaka, Bangladesh, and Concepción, Chile.

One of the main challenges in developing robust behavioural models, particularly in the Global South, is the absence of comprehensive and up-to-date mobility data. While conventional survey data can offer valuable sociodemographic insights, they often lack the necessary information for developing robust behavioural models, such as preferred departure time, arrival time, availability of mode, and travel time at other times of the day and for different mode. Moreover, emerging data sources like Microsoft Bing Maps and the Google Maps API, which offer reliable real-time mobility data in the Global North, often have more uncertainty and hence prove insufficient in the Global South. To complement travel diary survey data, there is increasing interest to utilise passive data sources such as mobile phone CDR, GSM, and GPS data. However, many models developed using passive data lack validation for policy implementation due to a lack of appropriate ground truth data.

To overcome these data challenges, this thesis investigates the application of emerging data sources to enhance the limited data available in the global south, particularly to develop advanced discrete choice and agent-based micro-simulation models. A new method is proposed to estimate travel time for alternative time periods using the Google Maps API and stated travel times. This estimation enables the development of mixed logit-based departure time choice models using survey based revealed preference (RP) data for car commuters in Dhaka, Bangladesh. The developed models facilitate the inference of required statistics, such as preferred departure time, sensitivity to travel time which are crucial for implementing appropriate strategies (e.g., peak spreading policy) to reduce traffic congestion.

In situations where travel time data for unchosen modes (e.g., public transport, auto-rickshaw, motorcycle) are unavailable, a congestion matrix is formulated based on car travel times obtained from Google Maps data. This matrix enables the development of a mode choice model using RP and SP data to generate travel demand for future scenarios, such as the implementation of Bus Rapid Transit (BRT) in Dhaka. Implementing this model in an agent-based simulation platform (MATSim) enables the simulation of potential BRT demand scenarios in the context of Dhaka, facilitating a more comprehensive assessment of future transport scenarios. Additionally, this thesis generates a sample of representative agents' activities, time choice, and movement trajectories on Dhaka's existing transportation system. The simulated scenario is then used to evaluate the accuracy of three mainstream passive data sources (i.e., GPS, CDR, and GSM data) in inferring spatiotemporal trajectory information such as stay location, travel

distance, and departure time. The results underscore the feasibility of leveraging emerging data sources for the development of advanced behavioural models. The framework can be extended in future for evaluating the implications of the margin of errors in estimating value of travel time savings and other welfare measures based on these passive data sources.

Furthermore, in order to better capture the behavioural dynamics (e.g., correlation, satiation, random heterogeneity) prevalent in the Global South, the thesis emphasises the importance of extending state-of-the-art models (i.e., departure time choice, activity choice, and time use decision). A comparison between the state-of-the-art models and other candidate models is carried out to examine the impacts of correlation, role of satiation, and activity start time on choice prediction. By employing a new polynomial functional form, the thesis investigates the correlation between outbound and duration choice, exploring its influence on time-of-day preference. Furthermore, it examines utility preferences that vary over time using time-dependent utility and upper bound time constraints based on activity start time, while jointly modelling activity type, start time, and duration. Finally, it examines the impact of satiation and correlation among alternatives on the prediction accuracy of activity duration using different activity models in the context of Concepcion, Chile. The results emphasise the importance of thoroughly examining behavioural dynamics within the context of the Global South and the need to improve and expand the existing state-of-the-art models to better capture the specific behavioural dynamics in the Global South.

The findings of the thesis are expected to be useful for the planners and policymakers in the Global South in the selection of appropriate data sources and modelling frameworks for improving the current state of practice in transport planning.

Acknowledgement

At first, I would like to thank the most to the Almighty for bestowing upon me the fortitude and patience necessary to complete this research successfully. I am pleased to express my sincere gratitude and hearty admiration to my honourable supervisors, Professor Charisma Choudhury and Professor Stephane Hess for their valuable suggestions and inspirations, constant guidance and support, and cordial encouragement towards the successful completion of this research. I owe much to Professor Charisma Choudhury for always being there when I needed her support, for continually monitoring my progress, giving outstanding feedback, and guiding me through my PhD studies. An immense thank to Professor Stephane Hess for assisting me in laying the groundwork for choice modelling from an urban planning perspective and opening the door for my entire PhD project. I am grateful to my both supervisor for accepting me as part of Choice Modelling Research group.

Without the funding support from the Schlumberger Foundation, Faculty for the Future Fellowship, this research project would not have been feasible. Their financial support for the female faculty members from developing nations inspired me and would inspire other female faculty members from those nations to pursue careers in the field of academic. Without the information we gathered from Chile and Bangladesh, the research would not be feasible. I appreciate the access to the data and the explanations provided by Mr. Anisur Rahman, Mr. Dhrubo Alam, and Dr. Jaun Antonio Carrasco.

I would also like to thank all the members from CMC research group for being a great bunch of people in and out of the research group. I must thank to Dr. David Palma and Dr. Panagiotis Tsoleridis who were always there for discussions about anything that I was unsure on related to magic package “apollo” in R. I am also grateful to Dr. Chiara Calastri, Dr. Thomas Hancock and Dr. Thijs Dekker for their support and words of encouragement during times when it was mostly needed. I would also like to extend my thanks to all the participants from the brownbag seminar, for their valuable suggestions in each of my brownbag session. Finally, I want to thank my peers Maximiliano Lizana Maldonado, Shuwei Lin, Prasetyaning Diah Rizky Lestari, Faza Bastarianto, Nur Karim for their friendship and the countless insightful discussions we had over dinners, phone calls and long walks.

I would like to express my sincere gratitude to my partner, Dr. Sarfaraz Gani Adnan for his unwavering support throughout my PhD journey. His love, patience, and encouragement were instrumental in helping me stay focused and motivated. I am also grateful to my parents (Late) Farida Yeasmin and Syed Nazmul Haque, for their unconditional love and support, and for always believing in me. Their guidance and wisdom have been invaluable to me, and I am honoured to be their child. Finally, I would like to thank my sister Dr. Nur E Zannat Fatema and brother-in-law Sakil Mostafa for their support and encouragement throughout my studies. They have inspired me with their unflinching faith in me. I am appreciative to my sibling Syed Mahmudul Haque Yamin for being my technological life support system throughout the journey. I feel honoured to refer to them as my family.

Table of Contents

List of figures

List of tables

List of supplementary figures

List of supplementary tables

List of abbreviations

ABM	Agent Based Models
ASC	Alternative Specific Constants
BRTA	Bangladesh Road Transport Authority
BRT	Bus Rapid Transit
CDR	Call Data Record
CNL	Cross-Nested Logit
DTCA	Dhaka Transport Coordination Agency
FL	Fractional Logit
FMNL	Fractional Multinomial Logit Model
FNL	Fractional Nested Logit Model
GDPR	General Data Protection Regulation
GEV	Generalised Extreme Value
GSM	Global System for Mobile Communication
IIA	Independence of Irrelevant Alternatives
ICT	Information and Communication Technology
LC FMNL	Latent Class Fractional Multinomial Logit Model
LC FNL	Latent Class Fractional Nested Logit Model
LC MDCEV	Latent Class Multiple Discrete Continuous Extreme Value Model
LC MDCNEV	Latent Class Multiple Discrete Continuous Nested Extreme Value Model
LL	Log-Likelihood
MMNL	Mixed Multinomial Logit
MNL	Multinomial Logit Model
MDCEV	Multiple Discrete Continuous Extreme Value Model
NL	Nested Logit
PAT	Preferred Arrival Time
PDT	Preferred Departure Time
PT	Public Transport
RAJUK	Rajdhani Unnayan Kartripakkha
RUM	Random Utility Model
RP	Revealed Preference
RMSE	Root Mean Square Error
SP	Stated Preference
SGT	Synthetic Ground Truth
TVSD	Time Value of Schedule Delay

Chapter 1 Introduction

1.1. Background

In the rapidly evolving urban landscapes of the Global South[1], transport planning faces unique challenges, particularly regarding data reliability and behavioural complexities (Falchetta et al., 2021). Data-related issues often arise due to fragmented data collection processes, limited resources, outdated infrastructure, and a lack of institutional capacity (Eros et al., 2014). Behavioural complexities can arise from rapid urbanisation, perpetual growth patterns, socio-economic diversity, and cultural nuances (Malayath and Verma, 2013; Pojani and Stead, 2018). Despite these challenges, there has been a limited emphasis on developing transport models and planning tools tailored to the Global South (Oviedo Hernandez and Nieto Combariza, 2021) in comparison with the decades of transport research conducted in the context of high-income countries in the Global North. As a result, researchers and policymakers often rely on transport models designed for the Global North. However, models originating from the Global North frequently overlook behavioural and contextual dynamics unique to the Global South, such as informal transport systems and distinct travel and activity patterns (Sabogal-Cardona et al., 2021; Thaithatkul et al., 2023). This oversight can lead to inaccuracies in the results. Further, the models developed in the context of the global north rely on high-quality and accurate data which are rare in the global south. Hence, applying a model framework developed for the global north to the global south context can introduce uncertainties in the results (Hasselwander and Bigotte, 2023). Therefore, it is crucial to develop transport models and tools focusing on the Global South while addressing both data challenges and behavioural complexities.

Informed transport planning requires comprehensive, reliable data to pinpoint transport needs, assess existing infrastructure efficacy, and prioritise transport project investments (Ortúzar et al., 2011; Harrison et al., 2020). Traditionally, data is manually collected through interviews (at roadside, on-board, or stations), household surveys, and telephone interviews (Amalan et al., 2023), providing valuable insights into respondents' sociodemographic characteristics as well as their actual (e.g., survey-based revealed preference (RP) data) and hypothetical travel preferences (e.g., stated preference (SP) data) (Devillaine et al., 2012; Helveston et al., 2018). Survey based RP data provides real-world market scenarios based on travellers' recent activities and choices and is widely used for behaviour studies and transport planning in the Global South. However, in the Global South, this is less frequently collected, resulting in a small sample size and being collected independently of census data, often with a focus on project-specific interests, like the construction of flyovers or subways (Ortúzar et al., 2011; Devillaine et al., 2012). This data collection approach has two main drawbacks. First, available household travel surveys often lack essential information needed by transport modellers to develop robust behavioural models (Enam and

[1] As of early 2022, the United Nations' Finance Center for South-South Cooperation listed 78 countries under the umbrella of Global South countries. This classification is not solely based on geographical location but rather encompasses nations facing common challenges. These challenges include unstable democracy, ongoing industrialisation processes, and a history of colonisation by the Global North. Additionally, populations within these countries are negatively affected by capitalist globalisation, poverty, population growth, war, disease, and border issues (WPR, 2023). It is estimated that more than two-thirds of the world's population will be residing in urban regions by the year 2050. However, 94% of the population growth between 2010 and 2015 was recorded in the global south, which was home to about 75% of the world's urban population in 2015. 27 out of the 33 megacities in the world are in the global south (Smit, 2021).

Choudhury, 2011; Li, Z. et al., 2020). For instance, developing a mode choice model requires information on available alternative modes and their level of service attributes such as travel time and cost. Similarly, departure time choice models need details about travel times at different time-of-the-day for various origins and destinations, as well as preferred departure or arrival times. Second, since small-scale household surveys do not align with the census data collection, applying advanced modelling frameworks such as microsimulation to assess the impact of specific plans before the implementation or predicting future demand under uncertainty becomes challenging (Zannat et al., 2023). The lack of connection between survey-based RP data and census complicates verifying the representativeness of small-scale data, emphasising the need for a large sample size. Consequently, assessing the applicability of models developed using survey-based RP data for the entire population, along with appropriate validation, becomes challenging. Additionally, these data often face econometric issues such as multicollinearity (correlation between observed features) and endogeneity (observed attributes by the agents but not by the modeller) (Börjesson, 2008). Therefore, a model developed solely relying on RP data collected by travel diary surveys may lack the necessary attributes that influence actual behaviour, hindering the development of a comprehensive understanding of behaviour essential for formulating effective transport policies. To address the limitation of survey-based RP data for model development, secondary data sources such as Microsoft Bing Maps and the Google Maps API offer reliable real-time mobility data in the Global North. However, in the Global South, these emerging sources often prove insufficient when used as standalone information providers. For example, in cities where public transport operates without fixed schedules or stoppage facilities, the Google Maps API fails to provide accurate vehicle travel information, public transport route details, or timing information (Zannat et al., 2022). To mitigate the limitations of survey-based RP data, another data source employed for behaviour modelling and transport planning is SP data. SP data has the potential to address issues related to multicollinearity and endogeneity by incorporating predefined attributes and alternative levels, which are not limited to currently available market products or attributes (Börjesson, 2009). Despite the benefits offered by SP data, its application in understanding complex or future behaviour, or for advanced modelling in transport planning in the Global South, remains limited. Only a few studies focusing on the Global South provide evidence of the application of SP data in understanding complex behavioural issues using advance modelling frameworks, particularly preferences for emerging transport modes (Enam, 2010; Venter, 2018). In contrast, countries in the Global North conduct frequent travel surveys (both RP and SP data) alongside censuses involving large representative samples of the entire population. This comprehensive approach helps them understand the movements of people and goods and the requirements for infrastructure development and policy formulation (Cornick et al., 2018; WESTAT, 2018).

Recently, the use of passively acquired data from the digital traces left by everyday technology, such as mobile phones (call data record (CDR), Global System for Mobile Communication (GSM)) and smartphone apps, Bluetooth, GPS, and smart cards, has become increasingly popular. This trend unveils new possibilities for innovative studies in transportation planning. These data sources offer several advantages, including up to date, near-real-time spatial and temporal information containing a significant amount of individual-level data that is more accurate, precise, and cost-effective (Xu and González, 2017; Tsumura et al., 2022; Amalan et al., 2023). Additionally, big data can be used to reconstruct large-scale trajectory data for larger sample sizes and longer observation periods (Shen and Stopher, 2014; Yue et al., 2014). Notwithstanding these advantages, the collection, processing, and analysis of big data can present various challenges (both computational and application) that must be addressed when using the data for transportation planning (Cortés et al., 2011; Kusakabe and Asakura, 2014; Shen and Stopher, 2014; Zhang, Y. et al., 2018). Utilising emerging data sources for transport planning in low-income countries from the Global South is particularly challenging due to a lack of required resources, technology, and reliable ground truth information. Overcoming these challenges, including data processing, computational issues, and result validation, becomes especially difficult in this context. Few studies from the Global South have utilised big data to analyse behavioural complexity and conduct transportation planning research (Iqbal et al., 2014; Munizaga et al., 2014; Bwambale, A. et al., 2019a; Bwambale, A. et al., 2019b; Bwambale, A. et al., 2019). However, the findings of these studies may have limited applicability in the policymaking process due to a lack of accuracy assessment and validation against appropriate ground truth data.

In addition to data issues, transport planning in the Global South faces challenges from prevailing behavioural dynamics that influence travel demand and corresponding choices such as mode choice, travel and activity patterns, and mobility preferences (Hasselwander and Bigotte, 2023). Significant differences in travel behaviour and demand patterns between the Global North and south exist, largely due to variations in socio-economic conditions, technological adoption, cultural distinctions, social norms, policy frameworks, and government interventions (Uteng and Lucas, 2018). In the Global South, lower income levels, high rates of informality (both in employment and transport services), and increased income inequality significantly influence travel decisions, shaping unique travel patterns (Pojani and Stead, 2018). For example, empirical investigations have shown that people tend to work longer hours when wage rates decrease (Niyogi, 2015). Additionally, certain activity duration requirements are intricately linked with other activity decisions and their associated dimensions, such as start time and activity location within a 24-hour budget constraint, including travel decisions such as the time of travel and mode selection from available alternatives (Eluru et al., 2010; Pinjari and Bhat, 2010; Golshani et al., 2018). Such intricate relationships among different activity-travel related dimensions are prevalent across diverse geographical contexts, including both the Global South and north. Considering these behavioural dynamics in travel demand modelling is essential for effectively designing and implementing transportation policies and infrastructure projects that cater to the needs and preferences of diverse populations. Over the last two decades, behavioural models have evolved to recognise complex human behaviour more accurately, aiming to estimate travel demand more precisely. Consequently, modelling paradigms have shifted from fixed peak hour demand models to peak spreading policy measures to optimise existing capacity (Loudon et al., 1988; Feil, 2010; Kreindler, 2023). Earlier studies focused on more tractable human travel behaviour, such as mode and route choice (Bhat, 1998; Pozsgay and Bhat, 2001). However, new transportation concerns necessitate further study into the latent and dynamic nature of human mobility preferences, including unobserved and latent preferences, correlation and endogeneity among activity-travel dimensions (Bekhor and Prashker, 2008; Vij and Walker, 2014; Prato et al., 2017; Guerrero et al., 2021). Despite advancements in modelling intended to capture the complexities of human travel behaviour, only a few demand and travel behaviour models have been developed addressing the prevailing behavioural complexities in the Global South (Le Loo et al., 2015). For example, the trip-based "four-step model", a first-generation demand model, was developed in the US in the late 1960s. While these four-step-models have been applied in the Global South, evidence indicates limited adjustment, use of proxy indicators, and ad-hoc assumptions. Hasan (2007) developed a four-step demand model in Dhaka, Bangladesh, by adapting a rule-based choice model for cars, assuming that travellers with cars will always choose cars regardless of situational constraints. In the Indian context, traffic assignment in demand forecasting often does not consider non-motorised modes (Guitink et al., 1994; Malayath and Verma, 2013). Furthermore, well-known activity based disaggregate models such as STARCHILD (simulation of travel/activity responses to complex household interactive logistic decisions), ALBATROSS (A learning-based transportation-oriented simulation system), TASHA (Toronto area scheduling model for household agents), CUSTOM (comprehensive utility-based system of activity-travel scheduling options modelling), CEMDAP (comprehensive econometric microstimulator for activity-travel patterns) (Bhat et al., 2004; Nurul Habib, 2018) are developed and implemented with a focus on countries in the Global North. Few studies have highlighted the data issues, modelling challenges, application, and validation complexities of these advanced modelling frameworks focusing on the Global South (Yagi and Mohammadian, 2010). Additionally, the development of travel demand models often requires various types of data (e.g., census, travel diary, land use, infrastructure, etc.) that are not readily available in Global South countries due to financial constraints. An economically viable solution to this issue could be to transfer the model output across time and space, either in its original state or with minimal updating. However, a review of research on transferability in the context of the Global South reveals that the models within countries in the Global South are either non-transferable or partially transferable in some instances, as the travel-related attitudes and behaviours are largely non-generalisable (Yagi et al., 2012; Sanko et al., 2014; Flavia and Choudhury, 2019). To expand our understanding of the behavioural dynamics, predominate in the Global South, more research is needed to address the limitations of the current literature on the region.

For effective transport planning in the Global South, it is crucial that transport models and planning tools can adapt to the increasing challenges presented by evolving transportation dynamics and emerging policy agendas (Wise et al., 2017). Real-world experiments, such as assessing the effectiveness and efficiency of constructing flyovers after the implementation, can be both financially burdensome and time-consuming (Jones, 1983). Therefore, conventional planning approaches often rely on demand forecasting models to monitor and anticipate shifts in demand. While both four-step demand model and activity-based forecasting models have been valuable tools for understanding travel behaviour and predicting demand (Rasouli and Timmermans, 2014), agent-based model (ABM) represents a new generation of transport models that offer several advantages and addresses limitations associated with traditional modelling approaches (Chen and Cheng, 2010; Axhausen et al., 2016; Kagho et al., 2020). For instance, when examining the impact of specific policies on the network level using an activity-based model, integrating demand generation with traffic assignment still requires a robust solution (Bastarianto et al., 2023). Moreover, many cities in the Global South rely on informal transport systems (e.g., human-hauler, motorcycle) and diverse access-egress mode (e.g., rickshaws along with walking and cycling). These systems often operate outside formal regulatory frameworks and exhibit complex dynamics that may not be adequately captured by traditional modelling approaches. On the other hand, ABM provides the flexibility to incorporate multiple attributes of agents and their environment in different layer formats and simulate the model to understand urban traffic flows (Grether et al., 2008; Manley et al., 2014;

Małecki, 2018), activity behaviours (Arentze et al., 2010; Märki et al., 2014; Čertický et al., 2015; Gkiotsalitis and Stathopoulos, 2015; Shabanpour et al., 2017), changes in land use and effects on environment (Zhang, S. and Zhao, 2018), performance assessment of service (Gao et al., 2016; Ji et al., 2018; Levine et al., 2018), accessibility of location (Huang, 2019), and location decision of housing (Ding et al., 2018). Despite the advantages of ABM in accurately capturing the multi-dimensionality of travel decisions, including spatial, social, and economic factors, as well as the underlying heterogeneity in travel behaviour resulting from diverse travel patterns among heterogeneous populations, very few case studies can be found in the literature that utilise ABM frameworks to address complex urban transport issues (Bossert et al., 2020; Novizayanti et al., 2021; de Assis et al., 2023). Among these limited case studies, only a few have highlighted the data related issues, behavioural complexities, validation, and alignment of these behavioural and travel demand modelling frameworks with existing policies (Salas-Rodríguez and Rivas-Tovar, 2022).

Therefore, to expand our understanding of the prevalent data issues and behavioural dynamics in the Global South, further research is needed to rectify the deficiencies evident in the current literature on the region by conducting thorough investigations into existing modelling frameworks and emerging advanced modelling approaches.

1.2. Research Gaps

This thesis attempts to address different data-related challenges associated with the use of conventional and emerging data sources to develop advanced choice models and microsimulation models. Additionally, it explores the potential of extending the state-of-the-art model to address the data issues and behavioural dynamics, specifically focusing on the Global South. Even though some of the data related issues and behavioural dynamics raised in this study may be relevant to the Global North, empirical investigation was done with an emphasis on the Global South because those issues are more pertinent to that region.

Research gap 1

The methodology used for transport modelling initiatives in the Global South has been shifting from aggregate modelling to disaggregate modelling, though at a very early stage (Yagi and Mohammadian, 2010; Miskeen and Rahmat, 2011; Manoj and Verma, 2017). State-of-the-art behavioural models, however, usually encounter difficulties in addressing the complex, dynamic, and ever-evolving transport development issues of the Global South due to a lack of reliable activity and travel data. Conventional data sources, such as RP data collected by manual travel surveys, commonly utilised for understanding activity and travel patterns in the Global South, often lack the information (e.g., data for unchosen alternatives, preferred departure time) needed to develop robust behavioural models. As a result, the use of unreliable data during the modelling stage can lead to inaccurate outcomes and suboptimal decisions (Gordon et al., 2013). The use of reliable data is crucial regardless of whether the behavioural or demand models are developed using conventional or emerging data. However, conducting a survey and manually gathering all the necessary information to construct a robust behavioural model, or solely relying on observed data to capture the influence of emerging factors, may not be practical (Weiner, 1990).

Studies highlight the potential of new and emerging data sources in enhancing transportation modelling and policy making (Harrison et al., 2020). For example, various emerging third-party data sources, such as Google Maps API data[2] provide valuable individual mobility information often lacking in the conventional data sources (Tsoleridis et al., 2022). A crucial element in the development of discrete choice models (e.g., mode, departure time, and route selection) for predicting user demand is the precise measurement of travel time for unchosen alternatives, an attribute lacking in RP data (Li, R. et al., 2013). In the absence of travel time data for unchosen alternatives in RP data, empirical evidence suggests using Google Maps data to extract travel times for various modes and times of the day (Bwambale, A. et al., 2019a; Tsoleridis et al., 2022). While the use of such sources has been proven to provide precise travel time information in the context of the Global North, studies focusing on the Global South remain limited, especially those exploring the application of emerging data sources in the development of advanced behavioural modelling. Furthermore, in addressing the challenges posed by unchosen alternatives, evidence from the Global South suggests employing SP data (though on a limited scale) or the use of a network model to enhance understanding of mode choice (Enam and Choudhury, 2011). Similar endeavours to construct departure time choice models encounter difficulties, primarily stemming from the need for extensive travel time data across various time frames and origin-destination pairs. This requirement often proves challenging to accommodate within an SP design due to the considerable number of alternative choices (Hess et al., 2007). The network model often proves insufficient in the context of the Global South as it does not accommodate informal modes of transport (Hasan, 2007; Malayath and Verma, 2013). Efforts to develop departure time choice models with either a limited range of coarse alternatives or a finer selection but over a restricted time period may fail to capture the comprehensive range of time-of-day choices necessary for understanding temporal demand patterns and designing effective peak spreading policies. This emphasises the need for research focusing on the Global South aimed at identifying and refining potential emerging sources to enhance the limited data available for advanced modelling. The first research question of this thesis is:

"What are the challenges associated with using emerging third-party data sources (e.g. Microsoft Bing, Google Maps API, etc.) in the context of the Global South and how can they be made suitable for improving the state-of-the-art models?"

Further, in more recent years, passively generated trajectory data (e.g., Global Positioning System (GPS), Call Detail Records (CDR), Global System for Mobile Communication (GSM), social media, public transport smart card records have been gaining popularity in transport modelling (Zannat and Choudhury, 2019). However, they are less commonly used for transport planning in the global south. One potential reason that commonly inhibits the wider use of emerging data for mobility research is the lack of study on the relative accuracy assessment of mobility related trajectory information extracted from these sources. Before using the extracted travel data from anonymous passive sources for modelling purposes, it is crucial to conduct an accuracy assessment. This evaluation is necessary to ascertain the degree of alignment between the extracted mobility information and the actual events or conditions it aims to represent. This is particularly important due to the presence of inherent noises within the data, which significantly affect both its positional and temporal accuracy. The main challenge to such evaluation

[2] The emerging third-party data sources refer to Microsoft Bing, and Google Maps API which enables the calculation of distances and travel times by several routes and modes of transportation, including transit, driving, walking, or bicycling.

and/or validation has been the lack of ground truth data, which is typically unavailable for privacy reasons (Munizaga et al., 2014). Future research is therefore required to systematically investigate the extent of inaccuracies of different types of passive trajectory data sources using appropriate ground truth information. Therefore, the second research question of this study is:

"How can modellers assess the accuracy of passive trajectory data in the absence of suitable ground truth information?"

Research Gap 2

There is debate over what constitutes a good model, including whether it improves understanding and aids in explaining behavioural complexity or whether it has higher prediction accuracy (Hensher and Button, 2000). There is consensus that the model needs to be behaviourally consistent. To accurately forecast travel and activity demand in the context of the Global South, it is important that activity and travel models can accommodate various activity and travel-related attributes, and their interactions, and capture potential determinants influencing decision-making in their choice (Yagi and Mohammadian, 2010). Some behavioural modelling studies concentrate on a single-choice decision, while others use a joint model to capture the complexity of behaviour (combining different choice dimensions). In the context of the Global South, modelling single or joint choice decisions based solely on experiences from the Global North may result in simplified representations of decisions due to limited observed attributes in the conventional data. For instance, the concurrent departure time choice model developed by Ben-Akiva and Abou-Zeid (2013) and Hess et al. (2005) using the multinomial logit model (MNL), a classical discrete choice model utilised for forecasting travel demand at different times of the day, overlooking the prevalent cultural factors in the Global South such as long working hours and strict office schedules leading to the association between departure time choices and duration requirements. Similarly, in modelling activity type and duration choice using the multiple discrete-continuous extreme value model (MDCEV), a common practice in the Global North is to consider satiation, implying diminishing marginal utility. However, assumptions about the existence of satiation for all activities may not be applicable in the Global South due to income inequality and poverty. In such contexts, individuals from low and middle-income demographics may perceive longer hours as beneficial rather than exhausting. Indeed, satiation is person and context specific (McSweeney and Swindell, 1999; Kapur and Bhat, 2007; Galak et al., 2011). Moreover, the typical approach to modelling activity type and duration choice involves aggregating various activities under common categories such as work and education etc. However, activities may need to be redefined based on their starting times, and activity models need to accommodate both the activity start time and duration choice dimensions within a unified framework. This necessity arises because these two decisions are closely related to each other, with common factors influencing both the choice of duration and start time (Golshani et al., 2018). This unified approach is particularly essential for the Global South, where the starting time of an activity often implies an opportunity for flexibility and reflects the obligations associated with the activities based on budget. Therefore, in the current state-of-the-art model, many complex real-life situations, including but not limited to the relationship between outbound departure time and activity duration, time-dependent choices, activity-wise time use constraints, and the effects of satiation in activity decisions, are yet to be implemented or extended. Moreover, the use of advanced behavioural models and their applications in comprehending behavioural dynamics is notable in the research and cases related to the Global North. While models designed with a focus on the Global North could be significant within the context of the Global Northern regions, it becomes crucial to account for the dominant behavioural dynamics specific to the Global South when contemplating behavioural models from that region.

"How can advanced models developed in the context of the Global North be adopted or extended to accommodate the complexities of travel behaviour in the Global South?"

Research Gap 3

The majority of researchers concentrated on either developing behavioural models — which offer rational ways to examine agents' behaviour — or assessing the performance of transportation systems (Ben-Akiva et al., 2002; Hess et al., 2004; Bhat, 2008; Ben-Akiva and Abou-Zeid, 2013; Naim et al., 2022). However, the effects of various transportation options on different socioeconomic groups are considerably more important to policymakers. Transportation impact assessment is challenging as it involves many uncertainties (Deng, 2013; Wang et al., 2018). The lack of future data and suitable behavioural models makes it more difficult to predict the influence of future transport infrastructure. In transportation planning, simulation models are developed to analyse uncertainty and numerically illustrate how the system functions under various alternative policy scenarios. Simulation models provide the opportunity to generate futuristic synthetic data and analyse the future scenario in a virtual system without implementing the measures in the real world (Kitamura and Kuwahara, 2006; Xiong et al., 2017; Kaddoura and Nagel, 2019). However, conventional trip-based simulation models are inadequate in capturing individual decision-making processes as they primarily focus on aggregated travel behaviour (Zhuge et al., 2014). While activity-based demand models address the limitations of trip-based models, distinguishing the impact of implemented policies spatially using activity-based demand generation and scheduling is often challenging. This difficulty arises from the fact that only a limited number of activity-based models incorporate route choice, thereby hindering the ability to accurately assess the spatial effects of implemented policies (Rasouli and Timmermans, 2014). This necessitates further integration of the activity-based demand model with traffic assignment, which still lacks robust solutions. The emergence of a new generation of simulation models, namely agent-based simulation models, offers a promising approach to replicating the complexity within a system. These models have been widely adopted in transport planning to address complex issues such as congestion and the impact of individual choice on infrastructure (Vovsha, 2019). However, many case studies employing agent-based modelling (ABM) frameworks in transport-related simulations have focused on countries in the Global North (Bastarianto et al., 2023). Only a few studies from the Global South have highlighted the data-related issues, implemented the behavioural models in the ABM simulation platform, and assessed the impact of certain policies without their implementation on the ground focusing on the Global South (Yagi and Mohammadian, 2010). Therefore, the third research question of this thesis is:

"How can complex transport policy scenarios be investigated using cutting-edge simulation models in the Global South?"

1.3. Aim and objectives

The overarching aim of this thesis is to contribute to travel behaviour research in the resolution of some of the unresolved data and modelling related issues from the Global South. To achieve this goal, three objectives were formulated.

Objective 1: To explore how emerging third-party data (e.g. Google Maps API) and passively generated mobility data (e.g. GPS, GSM, and CDR) can enhance the limited traditional data available in the Global South, to facilitate the application of advanced modelling frameworks.

> Specific objective 1.1: To develop a departure time choice model of car commuters by advanced discrete choice modelling.
>
> Specific objective 1.2: To investigate the relative precision of passive data for inferring spatiotemporal travel trajectories with appropriate ground truth.

Objective 2: To evaluate the existing state-of-the-art model in addressing emerging issues of transport planning dynamics in the context of the Global South.

> Specific objective 2.1: To develop a joint departure time choice model capturing correlations between departure times and activity duration.
>
> Specific objective 2.2: To investigate the intricate interplay among the choices of activity type, duration, and start time that influence decisions related to both mandatory and discretionary activities.
>
> Specific objective 2.3: To investigate the role of satiation in the activity type and time use decision.

Objective 3: To develop agent-based micro-simulation models for forecasting demand for Bus Rapid Transit (BRT) by integrating multiple data sources, including revealed and stated preference data.

The framework of the thesis and contribution to the aim of the research is presented in Figure 1-1.

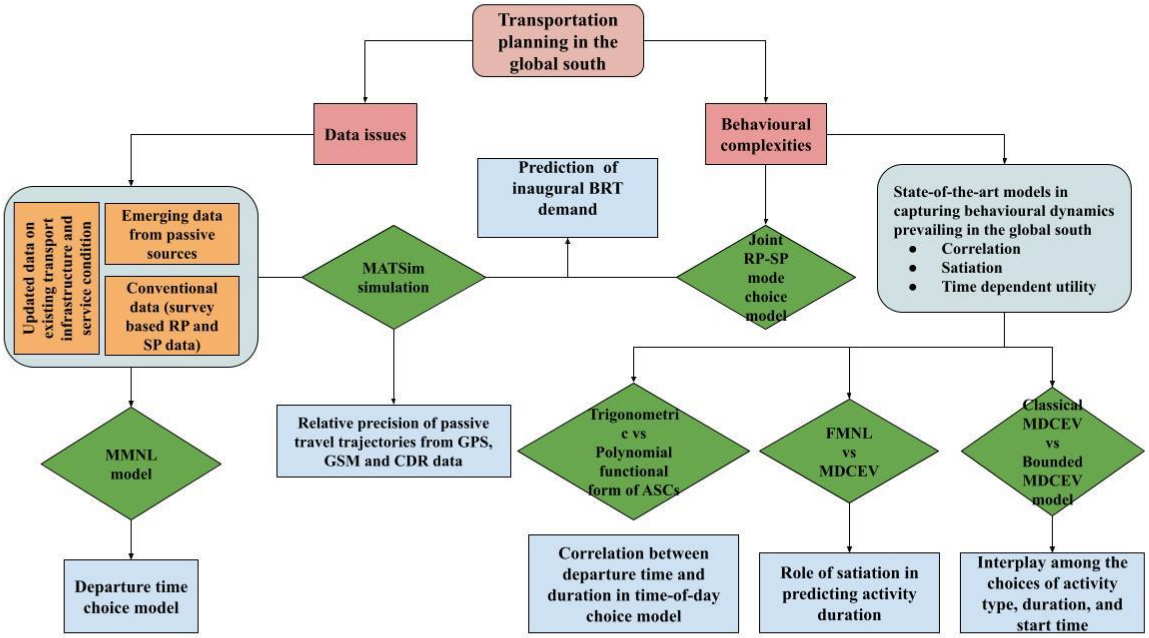


Figure 1-1 Thesis framework.

1.4. Thesis outline

The thesis is organised into eight chapters. The current chapter introduces the subject under study, the identified research gaps, and the overall objectives of the thesis. The remaining chapters analyse various methods that have been suggested to fill those gaps. The result from this thesis is summarised in the final chapter.

Chapter 2 presents a paper titled "Modelling departure time choice of car commuters in Dhaka, Bangladesh". Modelling departure time choice focusing on the Global South is very challenging due to the lack of dependable data sources. In this article, a new method was proposed to estimate the travel time for the full range of alternative time periods using Google Maps API and stated travel times when the Google Maps API is not deemed to be a reliable stand-alone source of travel time information. The purpose of this article was to take advantage of advanced discrete choice models to extend the state-of-the-art method for representing preferred departure time. In this article, instead of assuming a constant "preferred departure time" value for a specific market segment or a generic statistical distribution, the proposed model included two different statistical distributions for office workers and self-employed personnel acknowledging the high level of heterogeneity between and within each group. Furthermore, in this article, it was also quantified how the level-of-service attributes (e.g., travel time), socio-demographic characteristics (e.g., type of job, income, etc.), and situational constraints (e.g., schedule delay) affected car commuters' departure time choices.

Chapter 3 contributes to the further improvement of the departure time choice model. This chapter includes a paper titled "Modelling time-of-travel preferences capturing correlations between departure times and activity duration". In this article, a novel modelling framework was proposed to jointly estimate the departure and return times of travellers using a polynomial functional form of alternative specific constants (ASCs) with the goal of more accurately forecasting time-of-travel preferences. For a joint (outbound and return) departure time choice model, the suggested functional form captures the interaction between the constants of outbound departure time and duration. The proposed polynomial model was also contrasted with the state-of-the-art departure time choice model proposed by Ben-Akiva and Abou-Zeid (2013) who used a trigonometric functional representation.

Chapter 4 highlights the significance of time dimensions such as activity start time in an activity type and time use model using the mainstream activity modelling framework i.e., multiple discrete continuous extreme value model (MDCEV). The title of the article is “Joint modelling of activity type, start time and duration using a bounded multiple discrete-continuous framework: A case study of Dhaka, Bangladesh”. Using the bounded MDCEV model, activity type and duration models were estimated while accounting for utility differences based on activity start time. Additionally, time budget constraints based on activity start time (either starting the activity in the morning or afternoon) were accommodated while maintaining the total consumption budget 24-h. From a policy perspective, the proposed joint model using activity start time with activity type and duration choice provided rich situational information about activity and travel patterns which could be used for activity timing in an agent-based simulation model.

Chapter 5 raises the well-known debate over whether a good model should produce results with greater prediction accuracy or more behaviourally realistic outcomes. The title of the article is “The role of satiation in activity participation: a comparative analysis of fractional logit (FMNL) model and multiple discrete-continuous extreme value (MDCEV) model”. Data used for this work was sourced from Concepción, Chile[3]. Using different activity type and duration choice models, the impact of satiation and correlation were investigated in recovering the average time used for different activities. Furthermore, it explores the additional insights gained from incorporating random heterogeneity in time use models. Multiple discrete-continuous extreme value models (MDCEV) and the fractional logit (FL) models were used in the comparative analysis. The latent class MDCEV model and corresponding class profile were used to classify the heterogeneous sample into homogenous groups based on model fit and behavioural realism with respect to baseline preferences and satiation. Fractional nested logit models (FNL) and Multiple discrete continuous nested extreme value models (MDCNEV) were used to capture unobserved correlation among the alternatives. The findings from the investigation of the satiation effect could guide transport planners in predicting activity duration parsimoniously with greater prediction accuracy and formulate optimum market strategies that can address their travel needs.

Chapter 6 includes an article on “Developing an agent-based microsimulation for predicting the Bus Rapid Transit (BRT) demand in developing countries: A case study of Dhaka, Bangladesh”. The implemented Dhaka scenario integrated mode choice model, calibrated with multiple data sources (both RP and SP data), to realistically mimic the mode choice behaviour prevailing in the city. This is the first open scenario implemented for Dhaka, Bangladesh. Using the Dhaka scenario, different circumstances of inaugural BRT were simulated and the sensitivity of the outputs to different modelling assumptions (travel time, cost, and access mode) was tested. While direct model outputs presented in this article would be useful for the planners to maximise the ridership of the proposed BRT, the calibrated simulator would be also useful for the evaluation of other innovative transport modes in the context of Dhaka in the future.

Chapter 7 includes the article entitled “Investigating the relative precision of GPS, GSM, and CDR data for inferring spatiotemporal travel trajectories”. To compare the accuracy and measurement error of

[3] Chile is one of the 78 countries categorized as part of the Global South by the United Nations' Finance Center for South-South Cooperation (WPR, 2023).

passively collected trajectory data, this article suggested a simulation-based framework to circumvent the problem of the lack of trustworthy ground truth data. The statistical and spatial characteristics of individual mobility information (activity location, trajectory distance, departure time) extracted from the GPS, GSM, and CDR data were compared with the compatible ground truth data to assess the accuracy of passive trajectory data, both at the disaggregate and aggregate level. This article was an effort to address issues with the conventional ground truth data, including missing information (such as route choice), and imprecise spatial and temporal granularity (e.g., TAZ level location data).

Chapter 8 contains the discussion and findings from the earlier chapters' research and potential directions for future study are also referred.

1.5. References

Amalan, T.P., Du, B., Kumarage, A.S., Wickramasuriya, R. and Perez, P. 2023. High frequency data in Land use and transport integrated model: A review of sources and application. *Asian Transport Studies.* **9**, p100092.

Arentze, T.A., Ettema, D. and Timmermans, H.J. 2010. Incorporating time and income constraints in dynamic agent-based models of activity generation and time use: approach and illustration. *Transportation research part C: emerging technologies.* **18**(1), pp.71-83.

Axhausen, K.W., Horni, A. and Nagel, K. 2016. *The multi-agent transport simulation MATSim.* Ubiquity Press.

Bastarianto, F.F., Hancock, T.O., Choudhury, C.F. and Manley, E. 2023. Agent-based models in urban transportation: review, challenges, and opportunities. *European Transport Research Review.* **15**(1), p19.

Bekhor, S. and Prashker, J.N. 2008. GEV-based destination choice models that account for unobserved similarities among alternatives. *Transportation Research Part B: Methodological.* **42**(3), pp.243-262.

Ben-Akiva, M. and Abou-Zeid, M. 2013. Methodological issues in modelling time-of-travel preferences. *Transportmetrica A: Transport Science.* **9**(9), pp.846-859.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D. and Bunch, D.S. 2002. Hybrid choice models: Progress and challenges. *Marketing Letters.* **13**, pp.163-175.

Bhat, C.R. 1998. Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological.* **32**(6), pp.361-371.

Bhat, C.R. 2008. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological.* **42**(3), pp.274-303.

Börjesson, M. 2008. Joint RP–SP data in a mixed logit analysis of trip timing decisions. *Transportation Research Part E: Logistics and Transportation Review.* **44**(6), pp.1025-1038.

Börjesson, M. 2009. Modelling the preference for scheduled and unexpected delays. *Journal of Choice Modelling.* **2**(1), pp.29-50.

Bossert, A., Kersting, M., Timme, M., Schröder, M., Feki, A., Coetzee, J. and Schlüter, J. 2020. Limited containment options of COVID-19 outbreak revealed by regional agent-based simulations for South Africa. *arXiv preprint arXiv:2004.05513.*

Bwambale, A., Choudhury, C.F. and Hess, S. 2019a. Modelling departure time choice using mobile phone data. *Transportation research part A: policy and practice.* **130**, pp.424-439.

Bwambale, A., Choudhury, C.F. and Hess, S. 2019b. Modelling long-distance route choice using mobile phone call detail record data: a case study of Senegal. *Transportmetrica A: Transport Science.* **15**(2), pp.1543-1568.

Bwambale, A., Choudhury, C.F., Hess, S. and Iqbal, M.S. 2019. Getting the best of both worlds - a framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling. In: *15th World Conference on Transport Research India*

Čertický, M., Drchal, J., Cuchý, M. and Jakob, M. 2015. Fully agent-based simulation model of multimodal mobility in European cities. In: *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*: IEEE, pp.229-236.

Chen, B. and Cheng, H.H. 2010. A review of the applications of agent technology in traffic and transportation systems. *IEEE Transactions on intelligent transportation systems.* **11**(2), pp.485-497.

Cornick, P., Byron, C., Templeton, I. and Hurn, J. 2018. *National Travel Survey 2017 Technical Report.* Great Britain The Department for Transport.

Cortés, C.E., Gibson, J., Gschwender, A., Munizaga, M. and Zúñiga, M. 2011. Commercial bus speed diagnosis based on GPS-monitored data. *Transportation Research Part C: Emerging Technologies.* **19**(4), pp.695-707.

de Assis, R.F., Guerrini, F.M., Santa-Eulalia, L.A. and de Paula Ferreira, W. 2023. An agent-based model for regional market penetration of electric vehicles in Brazil. *Journal of Cleaner Production.* **421**, p138477.

Deng, T. 2013. Impacts of transport infrastructure on productivity and economic growth: Recent advances and research challenges. *Transport Reviews.* **33**(6), pp.686-699.

Devillaine, F., Munizaga, M. and Trépanier, M. 2012. Detection of activities of public transport users by analyzing smart card data. *Transportation Research Record.* **2276**(1), pp.48-55.

Ding, H., Yang, M., Wang, W. and Xu, C. 2018. Simulating and analyzing the effect on travel behavior of residential relocation and corresponding traffic demand management strategies. *Ksce Journal of Civil Engineering.* **22**(2), pp.837-849.

Eluru, N., Pinjari, A., Pendyala, R. and Bhat, C. 2010. An econometric multi-dimensional choice model of activity-travel behavior. *Transportation Letters.* **2**(4), pp.217-230.

Enam, A. 2010. *Developing a comprehensive mode choice model to capture the preferences for mass rapid transit in Dhaka.* . MSc. thesis, Bangladesh Universiy of Engineering and Technology.

Enam, A. and Choudhury, C.F. 2011. Methodological issues in developing mode choice models for Dhaka, Bangladesh. *Transportation research record.* **2239**(1), pp.84-92.

Eros, E., Mehndiratta, S., Zegras, C., Webb, K. and Ochoa, M.C. 2014. Applying the general transit feed Specification to the global South: Experiences in Mexico city, Mexico–-and beyond. *Transportation Research Record.* **2442**(1), pp.44-52.

Falchetta, G., Noussan, M. and Hammad, A. 2021. Comparing paratransit in seven major African cities: an accessibility and network analysis. *Journal of Transport Geography.* **94**, p103131.

Feil, M. 2010. *Choosing the daily schedule: Expanding activity based travel demand modelling*. thesis, ETH Zurich.

Flavia, A. and Choudhury, C. 2019. Temporal transferability of vehicle ownership models in the developing world: case study of Dhaka, Bangladesh. *Transportation Research Record.* **2673**(3), pp.722-732.

Galak, J., Kruger, J. and Loewenstein, G. 2011. Is variety the spice of life? It all depends on the rate of consumption. *Judgment and Decision making.* **6**(3), pp.230-238.

Gao, M., Zhou, L. and Chen, Y. 2016. An alternative approach for high speed railway carrying capacity calculation based on multiagent simulation. *Discrete Dynamics in Nature and Society.* **2016**.

Gkiotsalitis, K. and Stathopoulos, A. 2015. A utility-maximization model for retrieving users' willingness to travel for participating in activities from big-data. *Transportation Research Part C: Emerging Technologies.* **58**, pp.265-277.

Golshani, N., Shabanpour, R., Auld, J. and Mohammadian, A. 2018. Activity start time and duration: incorporating regret theory into joint discrete–continuous models. *Transportmetrica A: transport science.* **14**(9), pp.809-827.

Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H. and Attanucci, J.P. 2013. Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. *Transportation Research Record.* **2343**(1), pp.17-24.
Grether, D., Chen, Y., Rieser, M. and Nagel, K. 2008. Integrating a simple model for mode choice into a multi-agent simulation of travel behavior and traffic flow.
Guerrero, T.E., Guevara, C.A., Cherchi, E. and Ortúzar, J.d.D. 2021. Addressing endogeneity in strategic urban mode choice models. *Transportation.* **48**, pp.2081-2102.
Guitink, P., Holste, S. and Lebo, J. 1994. *Non-motorized transport: confronting poverty through affordable mobility.*
Harrison, G., Grant-Muller, S.M. and Hodgson, F.C. 2020. New and emerging data forms in transportation planning and policy: Opportunities and challenges for "Track and Trace" data. *Transportation Research Part C: Emerging Technologies.* **117**, p102672.
Hasan, S. 2007. Development of a travel demand model for Dhaka City.
Hasselwander, M. and Bigotte, J.F. 2023. Mobility as a Service (MaaS) in the Global South: research findings, gaps, and directions. *European Transport Research Review.* **15**(1), p27.
Helveston, J.P., Feit, E.M. and Michalek, J.J. 2018. Pooling stated and revealed preference data in the presence of RP endogeneity. *Transportation Research Part B: Methodological.* **109**, pp.70-89.
Hensher, D.A. and Button, K.J. 2000. *Handbook of transport modelling.*
Hess, S., Bierlaire, M. and Polak, J. 2004. *Estimation of value-of-time using Mixed Logit models.*
Hess, S., Daly, A., Rohr, C. and Hyman, G. 2007. On the development of time period and mode choice models for use in large scale modelling forecasting systems. *Transportation Research Part A: Policy and Practice.* **41**(9), pp.802-826.
Hess, S., Polak, J.W. and Bierlaire, M. 2005. Functional approximations to alternative-specific constants in time-period choice-modelling. In: *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction, Proceedings of the 16th International Symposium on Transportation and Traffic Theory*: Elsevier Amsterdam, pp.545-564.
Huang, R. 2019. Simulating individual work trips for transit-facilitated accessibility study. *Environment and Planning B: Urban Analytics and City Science.* **46**(1), pp.84-102.
Iqbal, M.S., Choudhury, C.F., Wang, P. and González, M.C. 2014. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies.* **40**, pp.63-74.
Ji, Y., Gao, L., Chen, D., Ma, X. and Zhang, R. 2018. How does a static measure influence passengers' boarding behaviors and bus dwell time? Simulated evidence from Nanjing bus stations. *Transportation Research Part A: Policy and Practice.* **110**, pp.13-25.
Jones, P.M. 1983. *A new approach to understanding travel behaviour and its implications for transportation planning*. thesis, University of London.
Kaddoura, I. and Nagel, K. 2019. Congestion pricing in a real-world oriented agent-based simulation context. *Research in Transportation Economics.* **74**, pp.40-51.
Kagho, G.O., Balac, M. and Axhausen, K.W. 2020. Agent-based models in transport planning: Current state, issues, and expectations. *Procedia Computer Science.* **170**, pp.726-732.
Kapur, A. and Bhat, C.R. 2007. Modeling adults' weekend day-time use by activity purpose and accompaniment arrangement. *Transportation Research Record.* **2021**(1), pp.18-27.
Kitamura, R. and Kuwahara, M. 2006. Simulation approaches in transportation analysis: Recent advances and challenges.
Kreindler, G. 2023. *Peak-hour road congestion pricing: Experimental evidence and equilibrium implications.* National Bureau of Economic Research.
Kusakabe, T. and Asakura, Y. 2014. Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C: Emerging Technologies.* **46**, pp.179-191.
Levine, J., Zellner, M., Arquero de Alarcón, M., Shiftan, Y. and Massey, D. 2018. The impact of automated transit, pedestrian, and bicycling facilities on urban travel patterns. *Transportation planning and technology.* **41**(5), pp.463-480.

Li, R., Chai, H. and Tang, J. 2013. Empirical study of travel time estimation and reliability. *Mathematical Problems in Engineering.* **2013**.

Li, Z., Hensher, D.A. and Ho, C. 2020. An empirical investigation of values of travel time savings from stated preference data and revealed preference data. *Transportation Letters.* **12**(3), pp.166-171.

Loudon, W.R., Ruiter, E.R. and Schlappi, M.L. 1988. *Predicting Peak-Spreading Under Congested Conditions.*

Malayath, M. and Verma, A. 2013. Activity based travel demand models as a tool for evaluating sustainable transportation policies. *Research in transportation economics.* **38**(1), pp.45-66.

Małecki, K. 2018. A computer simulation of traffic flow with on-street parking and drivers' behaviour based on cellular automata and a multi-agent system. *Journal of computational science.* **28**, pp.32-42.

Manley, E., Cheng, T., Penn, A. and Emmonds, A. 2014. A framework for simulating large-scale complex urban traffic dynamics through hybrid agent-based modelling. *Computers, Environment and Urban Systems.* **44**, pp.27-36.

Manoj, M. and Verma, A. 2017. A structural equation model based analysis of non-workers' activity-travel behaviour from a city of a developing country. *Transportation.* **44**, pp.241-269.

Märki, F., Charypar, D. and Axhausen, K.W. 2014. Agent-based model for continuous activity planning with an open planning horizon. *Transportation.* **41**(4), pp.905-922.

McSweeney, F.K. and Swindell, S. 1999. General-process theories of motivation revisited: the role of habituation. *Psychological bulletin.* **125**(4), p437.

Miskeen, M.A.A.B. and Rahmat, R.A.A.B.O. 2011. Development of disaggregate mode choice models of intercity travel in Libya. In: *2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference*: IEEE, pp.197-201.

Munizaga, M., Devillaine, F., Navarrete, C. and Silva, D. 2014. Validating travel behavior estimated from smartcard data. *Transportation Research Part C: Emerging Technologies.* **44**, pp.70-79.

Naim, M.N.H., Adnan, M.S.G., Dewan, A. and Zannat, K.E. 2022. Assessing the performance of public transport services in a developing country: A case study using data envelopment analysis. *Growth and Change.* **53**(1), pp.377-409.

Niyogi, R.K. 2015. *What to do, when to do it, how long to do it for: a normative microscopic approach to the labour leisure tradeoff*. thesis, UCL (University College London).

Novizayanti, D., Prasetio, E.A., Siallagan, M. and Santosa, S.P. 2021. Agent-based modeling framework for electric vehicle adoption transition in Indonesia. *World Electric Vehicle Journal.* **12**(2), p73.

Ortúzar, J.D.D., Armoogum, J., Madre, J.L. and Potier, F. 2011. Continuous mobility surveys: the state of practice. *Transport Reviews.* **31**(3), pp.293-312.

Oviedo Hernandez, D. and Nieto Combariza, M. 2021. Transport planning in the Global South. *International Encyclopedia of Transportation.* Elsevier, pp.118-123.

Pinjari, A.R. and Bhat, C. 2010. A multiple discrete–continuous nested extreme value (MDCNEV) model: Formulation and application to non-worker activity time-use and timing behavior on weekdays. *Transportation Research Part B: Methodological.* **44**(4), pp.562-583.

Pojani, D. and Stead, D. 2018. Policy design for sustainable urban transport in the global south. *Policy Design and Practice.* **1**(2), pp.90-102.

Pozsgay, M.A. and Bhat, C.R. 2001. Destination choice modeling for home-based recreational trips: analysis and implications for land use, transportation, and air quality planning. *Transportation research record.* **1777**(1), pp.47-54.

Prato, C.G., Halldórsdóttir, K. and Nielsen, O.A. 2017. Latent lifestyle and mode choice decisions when travelling short distances. *Transportation.* **44**, pp.1343-1363.

Rasouli, S. and Timmermans, H. 2014. Activity-based models of travel demand: promises, progress and prospects. *International Journal of Urban Sciences.* **18**(1), pp.31-60.

Sabogal-Cardona, O., Oviedo, D., Scholl, L., Crotte, A. and Bedoya-Maya, F. 2021. Not my usual trip: Ride-hailing characterization in Mexico City. *Travel Behaviour and Society.* **25**, pp.233-245.

Salas-Rodríguez, D. and Rivas-Tovar, L.A. 2022. Agent-Based Modeling for Evaluation of Transportation Mode Selection in the State of Guanajuato, Mexico. *Computación y Sistemas.* **26**(4), pp.1689-1701.

Sanko, N., Dissanayake, D., Kurauchi, S., Maesoba, H., Yamamoto, T. and Morikawa, T. 2014. Household car and motorcycle ownership in Bangkok and Kuala Lumpur in comparison with Nagoya. *Transportmetrica A: Transport Science.* **10**(3), pp.187-213.

Shabanpour, R., Golshani, N., Auld, J. and Mohammadian, A. 2017. Dynamics of activity time-of-day choice. *Transportation Research Record.* **2665**(1), pp.51-59.

Shen, L. and Stopher, P.R. 2014. Review of GPS travel survey and GPS data-processing methods. *Transport reviews.* **34**(3), pp.316-334.

Smit, W. 2021. Urbanization in the Global South. *Oxford Research Encyclopedia of Global Public Health.*

Thaithatkul, P., Chalermpong, S., Laosinwattana, W., Liang, J. and Kato, H. 2023. The relationship between RHA use and car purchases: Trends emerging in Bangkok, Thailand. *Asian Transport Studies.* **9**, p100095.

Tsoleridis, P., Choudhury, C.F. and Hess, S. 2022. Deriving transport appraisal values from emerging revealed preference data. *Transportation Research Part A: Policy and Practice.* **165**, pp.225-245.

Tsumura, Y., Asada, Y., Kanasugi, H., Arai, A., Shibasaki, R., Kawaguchi, H. and Yamada, K. 2022. Examining potentials and practical constraints of mobile phone data for improving transport planning in developing countries. *Asian Transport Studies.* **8**, p100043.

Uteng, T.P. and Lucas, K. 2018. *Urban mobilities in the global south.* Routledge London.

Venter, C.J. 2018. Segmenting the Market for New Modes using Stated and Revealed Preferences. *Transportation Research Record.* **2672**(8), pp.796-806.

Vij, A. and Walker, J.L. 2014. Preference endogeneity in discrete choice models. *Transportation Research Part B: Methodological.* **64**, pp.90-105.

Vovsha, P. 2019. Decision-Making Process Underlying Travel Behavior and Its Incorporation in Applied Travel Models. In: *Decision Economics. Designs, Models, and Techniques for Boundedly Rational Decisions 15*: Springer, pp.36-48.

Wang, L., Xue, X., Zhao, Z. and Wang, Z. 2018. The impacts of transportation infrastructure on sustainable development: Emerging trends and challenges. *International journal of environmental research and public health.* **15**(6), p1172.

Weiner, E. 1990. Transportation Data Used Currently for National Planning Activities. *Transportation Research Record.* **1271**, p13.

WESTAT. 2018. *2017 NHTS Data User Guide.* USA.

Wise, S., Crooks, A. and Batty, M. 2017. Transportation in agent-based urban modelling. In: *Agent Based Modelling of Urban Systems: First International Workshop, ABMUS 2016, Held in Conjunction with AAMAS, Singapore, Singapore, May 10, 2016, Revised, Selected, and Invited Papers 1*: Springer, pp.129-148.

WPR. 2023. *Global South Countries 2023.* [Online]. Available from: https://worldpopulationreview.com/country-rankings/global-south-countries

Xiong, C., Pan, Y., Lee, M., Zhu, Z. and Zhang, L. 2017. Integrated agent-based travel behavior and dynamic traffic microsimulation for ramp-metering analysis. *Transportation research record.* **2665**(1), pp.11-20.

Xu, Y. and González, M.C. 2017. Collective benefits in traffic during mega events via the use of information technologies. *Journal of The Royal Society Interface.* **14**(129), p20161041.

Yagi, S. and Mohammadian, A.K. 2010. An activity-based microsimulation model of travel demand in the Jakarta metropolitan area. *Journal of Choice Modelling.* **3**(1), pp.32-57.

Yagi, S., Nobel, D. and Kawaguchi, H. 2012. Time series comparison of models of auto and motorcycle ownership and mode choice in a changing transportation environment: Jakarta, Indonesia. *Transportation research record.* **2317**(1), pp.40-50.

Yue, Y., Lan, T., Yeh, A.G. and Li, Q.-Q. 2014. Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society*. **1**(2), pp.69-78.
Zannat, K.E. and Choudhury, C.F. 2019. Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science*. **99**(4), pp.601-619.
Zannat, K.E., Choudhury, C.F. and Hess, S. 2022. Modeling departure time choice of car commuters in Dhaka, Bangladesh. *Transportation research record*. **2676**(2), pp.247-262.
Zannat, K.E., Laudan, J., Choudhury, C.F. and Hess, S. 2023. Developing an agent-based microsimulation for predicting the bus Rapid transit (BRT) demand in developing countries: A case study of Dhaka, Bangladesh. *Transport Policy*.
Zhang, S. and Zhao, J. 2018. Low-carbon futures for Shenzhen's urban passenger transport: A human-based approach. *Transportation Research Part D: Transport and Environment*. **62**, pp.236-255.
Zhang, Y., Yao, E., Zhang, J. and Zheng, K. 2018. Estimating metro passengers' path choices by combining self-reported revealed preference and smart card data. *Transportation Research Part C: Emerging Technologies*. **92**, pp.76-89.
Zhuge, C., Shao, C., Gao, J., Meng, M. and Xu, W. 2014. An initial implementation of multiagent simulation of travel behavior for a medium-sized city in China. *Mathematical Problems in Engineering*. **2014**.

Chapter 2 Modelling departure time choice of car commuters in Dhaka, Bangladesh

Abstract

Dhaka, one of the fastest-growing megacities in the world, faces severe traffic congestion leading to a loss of 3.2 million business hours/day (WB, 2015). While peak-spreading policies hold the promise to reduce the traffic congestion levels, the absence of comprehensive data sources makes it extremely challenging to develop econometric models of departure time choices for Dhaka. This motivates this paper where we developed advanced discrete choice models of departure time choice of car commuters using secondary data sources and quantified how the level-of-service attributes (e.g., travel time), socio-demographic characteristics (e.g., type of job, income, etc.) and situational constraints (e.g., schedule delay) affect their choices. The trip diary data of commuters making home-to-work and work-to-home trips by personal car/ride-hailing services (957 and 934 respectively) was used in this regard. Given the discrepancy between the stated travel times and those extracted using the Google Maps API (direction API), a sub-model was developed first to derive more reliable estimates of travel time throughout the day. A mixed multinomial logit (MMNL) model and a simple multinomial logit (MNL) model were estimated for outbound and return trip respectively, to capture the heterogeneity associated with different departure time choice of car commuters. Estimation results indicated that the choices were significantly affected by the travel times, schedule delay and socio-demographic factors. The influence of the type of job on Preferred Departure Time (PDT) was estimated using two different distributions of PDT for the office employees and self-employed people (Johnson's S_B distribution and truncated normal respectively). In addition to being practically useful for devising peak-spreading policies in Dhaka, the proposed framework can be useful in other developing countries with similar data issues.

Keywords: Departure time choice, Discrete Choice Model, Schedule Delay, Developing country, Dhaka, Bangladesh

2.1. Introduction

Dhaka, the capital of Bangladesh is home to more than 15 million people. The population of Dhaka (within the RAJUK[4] jurisdiction area) is projected to be 26.3 million by the year 2035, predominantly due to migration from rural to urban areas (DTCA, 2015). To meet the mobility demand of this rapidly growing population, the number of motorised vehicles in the city, private cars in particular, is increasing at an alarming rate. According to the statistics of the Bangladesh Road Transport Authority (BRTA), a total of 145,000 private cars, 20,000 trucks, 14,000 buses and minibuses are currently registered in the Dhaka Metropolitan Area which is expected to grow at a rate of 34% annually. This escalating growth of the motorised vehicle coupled with the increasing usage of private vehicles is associated with the severe traffic congestion that cripples the city and results in a loss of 3.2 million business hours per day (Siddique et al., 2014).

The traffic congestion levels in Dhaka are worst during the morning and evening peak hours. According to the World Bank (WB, 2015), the average speed on Dhaka's urban network during the peak hours is approximately 8.75km/h, indicating that the travel time during the peak hours is almost triple the travel time during the off-peak hours. A large share of the trips made during these periods is presumed to be mandatory trips by commuters that are often hard to cancel or reschedule. However, different types of commuters are expected to have different levels of flexibility in their start times at the workplaces as well as commitments at home. It is, therefore, crucial to develop an understanding of the factors affecting the departure time choice decisions made by the commuters and how it varies with the type of their job and socio-demographic characteristics. It then boils down to identifying the appropriate modelling specifications that can reflect the behaviour of decision-makers about departure time choice that can be used to design policies to flatten the peak demand.

Although departure time choice is a crucial determinant in measuring the temporal and spatial distribution of travel demand (Dong and Cirillo, 2020), it has received less attention than mode or route choice (Hendrickson and Plank, 1984; Arellana et al., 2012). For example, previous travel demand models developed focusing on Dhaka city discussed methodological issues in developing the mode choice model (Enam and Choudhury, 2011) but there has been limited research done in the context of departure time choice. Key challenges to develop departure time choice model in the case of Dhaka (as well as many countries in the developing world) include the following:

(1) lack of dependable data sources to calculate the travel time for different origin-destination pairs.

(2) different opening and closing times of different types of institutions making it harder to infer the preferred departure time choice.

In the context of developed countries[5], several modelling approaches have been used to model departure time choice. Bhat and Steed (2002) developed a continuous time choice model for urban shopping trips. In parallel, a large number of studies have used discrete choice models to investigate departure time choice by dividing the continuous departure time variable into a finite set of discrete intervals. For example, Small (1982), McCafferty and Hall (1982), Hendrickson and Plank (1984), Holyoak (2008)

[4] Rajdhani Unnayan Kartripakkha (RAJUK) is the Capital Development Authority of the Government of Bangladesh

[5] Developed countries are the sovereign state with a very high Human Development Index (HDI) rating. World Bank uses a threshold for the classification between developed and developing country i.e., per capita income level of US$6,000 in 1987 prices (Fantom and Serajuddin, 2016).

used simple multinomial logit (MNL) structures to model departure time choices of commuters. MNL models have also been used to model time of day choice in the context of trips made during weekends and holidays (Yang et al., 2008; Chaichannawatik et al., 2019). In addition to the single facet model, Bhat (1998) used a joint MNL model and ordered generalised extreme value formulation for integrated models of mode and departure time choices. The study however was focused only on non-commute trips. De Jong et al. (2003) and Hess et al. (2007) used mixed multinomial logit (MMNL) models to capture influences of unobserved factors in the context of mode and departure time choices. However, all these studies and their applications were focused on countries in North America and Europe which have a very different socio-economic compositions (e.g., income and age distribution, gender roles, household size and family structure, etc.), work-culture (e.g., inflexible working hours, recording of arrival time at workplace, etc.), state of technological advancement (e.g., reliable internet access and uninterrupted power supply to work from home) and transport landscape (e.g. car-ownership levels, public transport accessibility, paratransit, etc.) compared to the developing countries. All these lead to significant differences in activity and travel behaviour (Dimitriou and Gakenheimer, 2011; Enam and Choudhury, 2011; Cervero, 2013) and affect the transferability of the models (Dimitriou and Gakenheimer, 2011; Cervero, 2013; Zannat et al., 2021). Further, modelling frameworks formulated for the developed countries are very often not directly applicable in the context of developing countries where detailed socio-demographic information and fine-scale spatial and temporal data are not available.

Very few studies have discussed the methodological issues and data challenges of modelling departure time choice in the context of developing countries (Anwar, 2012; Arellana et al., 2013; Zou et al., 2016; Li et al., 2018). Among these limited studies, Anwar (2012) proposed a departure time choice model in the context of Dhaka using primary data (RP) with a pre-defined classification of timeslots within a narrow range (7:30 – 8:50). The collected data was used to develop ordered logit models of departure time choice. However, the study focused only on officials who had office hours from 9:00 to 17:00 ignoring the rest of the working population. Further, the travel times used in calibrating the model lacked adequate temporal and spatial granularity. Other studies relied on hypothetical scenarios generated by stated preference (SP) survey data (Arellana et al., 2013; Li et al., 2018) or jointly used both RP and SP data in the modelling framework (Zou et al., 2016).

This motivates this research where we developed advanced discrete choice models of departure time choice of car commuters using secondary data sources. We proposed approaches to account for the data limitations and quantify how the level-of-service attributes (e.g., travel time), socio-demographic characteristics (e.g., type of job, income etc.) and situational constraints (e.g., schedule delay, activity duration) affect the departure time choices of different types of commuters. Trip diary data of commuters making home-to-work and work-to-home trips by personal car/ride-hailing services (957 and 934 respectively) was used in this regard. It may be noted that though SP data have been used in some of the departure choice modelling studies (De Jong et al., 2003; Hess et al., 2007; Arellana et al., 2012; Arellana et al., 2013; Thorhauge et al., 2019), it is prone to hypothetical bias and behavioural incongruence (Bwambale et al., 2019) and hence revealed preference (RP) data was deemed to be the better option. In this study, we highlighted the key challenges and methodological issues to model departure time choice using RP data from a developing country and proposed ways to address these issues.

The rest of the paper is organised as follows: the next section describes the data sources used in this study. The modelling issues are presented next followed by the description of the model structure and the estimation results. The findings are summarised in the end along with directions for future research.

2.2. Data

In comparison to other household travel surveys that are currently available for the selected case such as DTCB (2005) and JICA (2010) in this study, we used the travel diary survey data collected for the Dhaka Metropolitan Region (RAJUK area) by TYPSA[6] (https://www.typsa.com/en/) as part of the feasibility study and preliminary design for the constriction of Dhaka Subway, Bangladesh. The travel diary data from TYPSA represents a recently available data source. The decision to prioritise this dataset over others was driven by two crucial reasons. Firstly, the subway feasibility study considered the mobility demands of all transportation users, encompassing both motorised and non-motorised, as well as private and public/para transport users of the RAJUK area. This inclusive approach facilitates in identification of major high-density corridors for the RAJUK area, which combines an urban agglomeration with zones not necessarily urban in character but closely bound to the centre by employment or activities (RAJUK, 2021). Also, such an approach provides a holistic mobility profile for the greater Dhaka region. In contrast, alternative data sources were (DTCB, 2005; JICA, 2010) limited to the Dhaka City Corporation area. Secondly, the TYPSA dataset incorporated all newly available modes of transport currently operational in Dhaka city, including ride-hailing services such as Uber and Pathao for both cars and motorcycles. This ensures that the selected database has the potential to capture the contemporary transport landscape and mobility behaviour in the selected study area.

The survey was conducted from Monday to Saturday[7] between 28th February 2019 to 4th May 2019. A total of 35,000 households were surveyed in the RAJUK Area. A stratified sampling procedure with proportional allocation was applied to determine the number of households that will be surveyed in each sub-divided area. About 25,000 households were surveyed in the Dhaka City Corporation area, with the remaining 10,000 household interviews conducted in the rest of the RAJUK area. During the surveys, each household member was asked about trips made during the previous working day (from Sunday to Thursday). The questionnaire survey was divided into two parts: the first part focused on general household information (e.g., age, gender, education, occupation, income, car ownership) and the second part focused on trip-related information (e.g., departure time, travel mode, travel time) for each household member who has made at least one trip on the previous working day. Very short trips (less than 10 minutes of walking distance) and trips made by children (under 6 years) were not recorded. The survey was well planned to avoid trips made on Fridays, Saturdays, public holidays, hartal (strike days), election days, major events (like Ijtema), and during Ramadan. Before the survey in each zone, the survey correspondents communicated with local representatives including the ward commissioners, in order to gain approval of, and assistance for, the surveys. The 35,000-household travel diary survey data included 1,37,760 trip information including passenger trips by buses and human-haulers (public/para-transport available in Dhaka), non-motorised transport (walking, rickshaws, and biking), car (including ride-hailing

[6] TYPSA is a private consulting and engineering service group that provides consultancy services for carrying out feasibility studies and design for the construction of the Dhaka subway project.

[7] Friday is the generic weekly holiday in Bangladesh. Educational institutes and some offices are closed on Friday and Saturday.

car usage), auto-rickshaws, and motorcycles (including ride-hailing motorcycle usage). In the full dataset, there were 4,003 trips that involved the use of cars. The focus of this study was on work-related home-based trips commuted by car. Out of the 4,003 observed car trips, 1,217 were specifically for work purposes. Among these car commuting trips, this study included a total of 957 unique home-to-work trips and 934 unique work-to-home trips[8] that met the screening criteria. Therefore, our selected sample included observations from respondents, both with households that have cars and those without cars. Moreover, it is essential to highlight that the selected sample exhibited a predominant representation of individuals falling within the working age group of 26 to 60, comprising approximately 85% of the sample. Nevertheless, it is pertinent to note that the age range within the sample demonstrated variability, spanning from 17 to 83 years old, owing to the presence of part-time private employment opportunities and post-retirement engagement. Commuting trips with origins outside Dhaka have not been considered. The socio-demographic characteristics of the commuters are summarised in Table 2-1.

Table 2-1 Summary of socio-demographic characteristics of the commuters in the sample.

	Percentage	**Percentage**
	Home-to-work trip respondents (n=957)	**Work-to-home trip respondents (n=934)**
Gender		
Male	83.28	82.87
Female	16.72	17.13
Age		
<26	3.97	4.18
26 -40	37.93	38.00
41-60	47.23	47.43
>60	10.87	10.39
Monthly income		
<10,000 BDT *	1.07	0.87
10,000-20,000 BDT	4.06	3.71
20,001-30,000 BDT	5.34	5.35
30,001-40,000 BDT	9.71	9.72
40,001-60,000 BDT	17.93	18.12
>60,000 BDT	61.89	62.23
Level of education		
Below primary (Year 5)	3.66	3.64
Years 6-10	5.96	5.89
Secondary School Certificate	6.06	5.78
Higher Secondary School	9.51	9.64
Certificate	17.76	17.77
Bachelors	54.75	54.93
Masters	2.30	2.35
Others		
Occupation		

[8] There is an imbalance in the number of home-to-work and work-to-home commute trips by car as some travellers have used public transport and non-motorised modes in one of the legs of the trip.

Employed in public services	20.69	21.09
Employed in private jobs	35.63	35.44
Self-employed	43.68	43.47
Car ownership rate		
Do not have a car	12.54	7.50
Have at least one car	87.46	92.5

*1 BDT= 0.012 USD

It may be noted that though the original sample was representative of the population of Dhaka city, the sample used in the departure time choice models of this paper is expected to be biased towards high income and educated segments of the population as it focuses only on car users (who have higher affordability than the others).

2.3. Modelling issues

2.3.1. Choice Set Specification

The number and length (i.e. duration) of alternative time periods play an important role in the computation, interpretation and transferability of the departure time choice models (Ben-Akiva and Abou-Zeid, 2013). In a usual specification, a separate alternative specific constant (ASC) is recommended for each possible combination of home to work (outbound) and work to home (inbound) time period to capture the unexplained time preference of travellers. However, this can lead to a compounding problem of higher computational cost and complex parameter identification issue. For example, using 1-hour time periods ($N=24$) would lead to a requirement for 300 constants (following the rule $N(N+1)/2$), of which 299 ($N(N+1)/2-1$) can be estimated (Hess et al., 2007). To reduce this computational cost, we used a separate set of alternatives for the outbound (home-to-work) and return (work-to-home) trips. In Dhaka, different public and private organisation adhere to varying opening and closing times, depending upon the type of institution. For example, educational institutions typically start activities at 7:00 and end at 17:00, though evening educational institutions may operate on a different schedule (Monolina and Ekram, nd). Public/private offices generally follow a timeframe between 9:00 to 18:00, with variations for educational institutions or hospitals (Hashi et al., 2015; Abedin et al., 2020). Shopping centres typically open at 10:00 and close at 22:00 (16:00 to 19:00 peak hour), yet owners of shops or restaurants may opt for distinct operating hours based on their preferences (Bangladesh, 2023). Furthermore, the existence of flexibility and job protocols introduces variation in the choices made within the same occupation group. For instance, a doctor may engage in morning hospital duties while concurrently participating in private practice during designated hours in a private chamber. Taking into account the variations in the opening and closing times across different institutions, in this study, the choice set for outbound and return trips was carefully designed based on the observed distribution of the time-of-day choices (Figure 2-2 (a) and Figure 2-3 (a)). The choice set for the outbound (home-to-work) trips was assumed to range between 06:00-18:00. Between 06:00-12:00, we used 1-hour intervals (since the majority of the outbound trips were observed to be made before 12:00) and 2-hour intervals were used for the rest. The choice set for the return (work-to-home) trips was assumed to range between 10:00-24:00. Between 16:00-20:00, we used 1-hour intervals (since the majority of the return trips were observed to be made before 20:00) and 2-hour intervals were used for the rest. In the time choice model, the off-peak hours (6:00 – 7:00, 10:00 – 12:00) were considered as base alternatives.

2.3.2. Calculation of Travel Time

Departure time choice models require the calculation of travel times between origins and destinations for the chosen and unchosen time periods. In many cities, Google Maps, Open Street Maps, etc. provide reliable travel times for each alternative time period with sufficient spatial and temporal granularity. Examples include departure time choice models developed in the context of the Global North (Bwambale et al., 2019; Dong and Cirillo, 2020) where Google Maps Distance Matrix API/Direction API was used for deriving travel times during different time periods for different origin-destination pairs. Since Google Maps uses historical data to predict future traffic, it is considered that future traffic conditions based on Google Maps are more stable and represent better trends in traffic than real-time information. However, in the context of a developing country, it is hard to infer travel time accurately from Google Maps. For example, in Dhaka, both motorised and non-motorised vehicles share a common right-of-way making the travel times very sensitive to the proportion of different types of vehicles. Further, in Dhaka, traffic intersections are manually operated by traffic police which also makes it harder to reliably infer travel times between a specific origin-destination pair. Moreover, there are multiple types of public transport and paratransit services (e.g., human hauler, 'tempo's etc.), which tend to allow passengers to board and/or alight at almost any place. These also make it almost impossible to reliably predict travel times across the network. Finally, though Google Maps API can show the shortest path in Dhaka, the use of navigation technology is not widespread among car users. A major share of the cars is chauffeur-driven, and the chauffeurs use their intuition to select the route to travel instead of choosing the quickest or shortest path that would have been recommended by a navigation device. Using the best guess[9] traffic model from the Google Maps API, which provides the most accurate estimated travel time in traffic at different time of the day, the estimated travel time did not precisely align with the travel time stated by the users during the chosen time period. Consequently, we investigated other traffic models (such as pessimistic[10] and optimistic[11]) from the Google Maps (direction) API when calculating time in traffic at different time of the day. The comparison between the stated travel time (only available for the chosen time of travel) and the three different predicted travel times using different traffic models from Google Maps API showed variations in fit depending on the time of the day and origin-destination pair. This prompted us to estimate a sub-model to establish a relationship between the stated travel time and the best guess, pessimistic and optimistic travel times from Google Maps. Employing the sub-model, we determined the weight assigned to the travel time provided by each traffic model from Google Maps for both the chosen and unchosen time periods.

The proposed relationship among the stated travel time and predicted travel time using models from direction API can be expressed as follows:

[9] Best guess model returns the duration in traffic using both historical traffic conditions and live traffic. Live traffic becomes more important the closer the departure time is to now.

[10] Pessimistic model returns the duration in traffic, usually that should be longer than the actual travel time on most days, though occasional days with particularly bad traffic conditions may exceed this value.

[11] Optimistic model returns the duration in traffic, usually that should be shorter than the actual travel time on most days, however, occasional days often with a good traffic condition could be faster than this value.

$$T_{stated\ travel\ time\_i} = W_{i,1}T_{Best\ guess\_i} + W_{i,2}T_{Optimistic\_i} + W_{i,3}T_{Pessimistic\_i} + \varepsilon \quad (1)$$

Where, *i* is the alternative time period ($i \epsilon n$, where *n* is 7 for home to work and 5 for work to home trip)
$T_{stated\ travel\ time\_i}$ = Stated travel time by the respondents at the alternative time period *i*
$T_{Best\ guess\_i}$ = Measured travel time using best guess model of Google Maps API at the time period *i*
$T_{Optimistic\_i}$ = Measured travel time using optimistic model of Google Maps API at the time period *i*
$T_{Pessimistic\_i}$ = Measured travel time using pessimistic model of Google Maps API at the time period *i*
$W_{i,t}$ indicates the weights of measured travel time by different Google Maps models (*t*) for time period *i*, ε represents the error. $W_{i,1}$ and $W_{i,2}$ were estimated keeping $W_{i,3}$ as a reference point (with their sum fixed to 1). The relationship among $W_{i,t}$ and estimated parameters for different models at the different time periods can be expressed as:

$$W_{i,t} = \frac{e^{\beta_{i,t}}}{e^{\beta_{i,1}} + e^{\beta_{i,2}} + e^{\beta_{i,3}}} \quad (2)$$

Following the normal distribution, equation 1 and equation 2 were used to estimate $\beta_{i,t}$ for three Google Maps models. Weights calculated for different Google Maps models (*t*) are presented in Table 2-2. The network travel time for all alternatives was estimated using the Equation 3.

$$T_{network\ travel\ time\_i} = \hat{W}_{i,1}T_{Best\ guess\_i} + \hat{W}_{i,2}T_{Optimistic\_i} + \hat{W}_{i,3}T_{Pessimistic\_i} \quad (3)$$

Table 2-2 Calculated weights of different models used for travel time calculation.
(a) Home to Work Trip

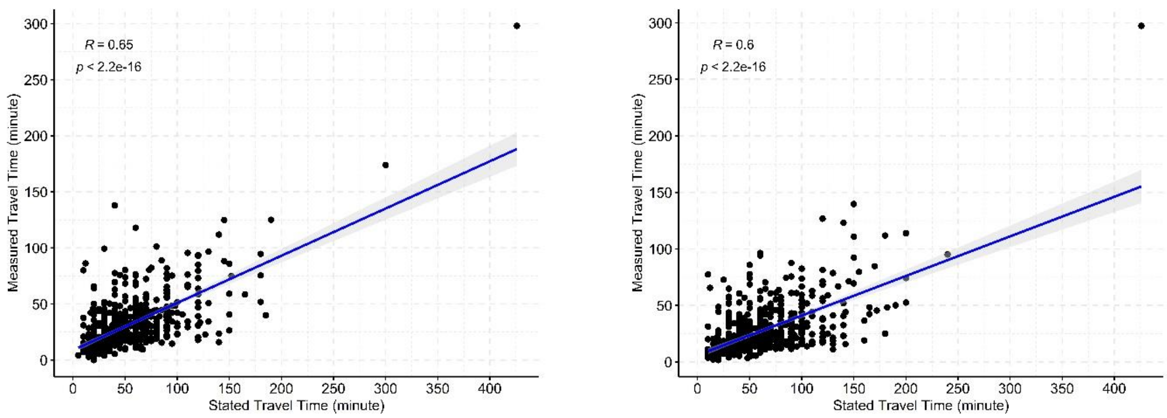
Departure Time	**Google Maps Model**	**Estimates ($\beta_{i,t}$)**	**Exp ($\beta_{i,t}$)**	**Weight ($\hat{W}_{i,t}$)**
		$\sigma = 31.4626$		
Before 7:00	Best Guess	14.2567	1554555.173	0.99958522[a]
	Optimistic	6.4678	644.0652237	0.000414137
	Pessimistic	0	1	6.43004E-07
		$\sum$	1555200.238	1
7:00 – 8:00	Best Guess	-15.9892	1.13757E-07	2.63623E-08
	Optimistic	1.1985	3.315140481	0.768257814
	Pessimistic	0	1	0.23174216
		$\sum$	4.315140594	1
8:00 – 9:00	Best Guess	-12.567	3.48515E-06	1.53532E-06
	Optimistic	0.239	1.269978537	0.559466341
	Pessimistic	0	1	0.440532123
		$\sum$	2.269982022	1
9:00 –10:00	Best Guess	-12.567	3.48515E-06	1.53532E-06
	Optimistic	0.239	1.269978537	0.559466341
	Pessimistic	0	1	0.440532123
		$\sum$	2.269982022	1

Departure Time	**Google Maps Model**	**Estimates ($\beta_{i,t}$)**	**Exp ($\beta_{i,t}$)**	**Weight ($\hat{W}_{i,t}$)**
10:00 – 11:00	Best Guess	-1.8409	0.158674555	0.055933598
	Optimistic	0.5177	1.678163432	0.591561252
	Pessimistic	0	1	0.35250515
		$\sum$	2.836837986	1
11:00 – 12:00	Best Guess	0.0022	1.002202422	0.499689173
	Optimistic	-5.6696	0.003449245	0.001719763
	Pessimistic	0	1	0.498591065
		$\sum$	2.005651666	1
12:00 and after 12:00	Best Guess	-0.7994	0.449598642	1.76382E-08
	Optimistic	17.0538	25490082.17	0.999999943[(b)]
	Pessimistic	0	1	3.92309E-08
		$\sum$	25490083.62	1

(b) Work to Home Trip

Departure Time	**Google Maps Model**	**Estimates ($\beta_{i,t}$)**	**Exp ($\beta_{i,t}$)**	**Weight ($\hat{W}_{i,t}$)**
10:00 – 12:00	Best Guess	0.0022	1.002202422	0.499689173
	Optimistic	-5.6696	0.003449245	0.001719763
	Pessimistic	0	1	0.498591065
		$\sum$	2.005651666	1
12:00 – 14:00	Best Guess	1.763	5.829900889	0.000898438
	Optimistic	8.7768	6482.101226	0.998947454[(c)]
	Pessimistic	0	1	0.000154109
		$\sum$	6488.931127	1
16:00 – 18:00	Best Guess	-9.1197	0.000109488	9.28015E-05
	Optimistic	-1.7165	0.179693977	0.1523084
	Pessimistic	0	1	0.847598799
		$\sum$	1.179803465	1
18:00 – 19:00	Best Guess	-5.1375	0.005872352	0.000662464
	Optimistic	2.0616	7.85853341	0.886526815
	Pessimistic	0	1	0.11281072
		$\sum$	8.864405762	1
19:00 – 24:00	Best Guess	7.1876	1322.924374	0.98801963
	Optimistic	2.7108	15.04130375	0.011233524
	Pessimistic	0	1	0.000746845
		$\sum$	1338.965678	1

Note: Given the close to 1 weight, only the Best Guess model has been used in (a) and only the Optimistic model has been used in (b) and (c) instead of the weighted models.



(a): Home-to-work trip (b): Work-to-home trip

Figure 2-1 Correlation between stated and measured travel time using weights calculated for different traffic models from Google Maps API.

Further, the fit of the results was tested by comparing the measured travel times (calculated using the estimated weights) with the stated travel times for each origin-destination pair (Figure 2-1). The correlation coefficients (0.65 and 0.6 for home-to-work and work-to-home trips respectively) signified the substantial positive association between the estimated and measured travel times.

2.3.3. Accounting for Schedule Delay

Schedule delay, which captures the disutility caused by travelling at times other than the desired time of travel, is a key variable in modelling departure time choice. Usually, the actual departure and travel times are recorded in RP surveys and the preferred arrival and preferred departure times (PAT and PDT) are missing in the data source. Asking direct questions to extract the information can also be biased due to potential subjective justification towards the actual or the intended arrival time (i.e., respondents may try to justify to themselves and/or the interviewer their actual behaviour is the optimum). Different studies have used different modelling approaches to model schedule delay. Koppelman et al. (2008) assumed that preferred departure time follows the same trend as the observed departure time. Though this assumption could be realistic for the air traveller, for regular commuters, it could be stringent. Ben-Akiva and Abou-Zeid (2013) suggested two methods: 1) assumption of a constant desired time of travel by market segment as PDT, and 2) assumption of a latent desired time of travel assuming a probability density function for the latent (unobserved) PDT. While these methods yielded good results in cities with homogeneous starting times of offices and businesses, the situation in Dhaka (as well as many other countries in the Global South) is more complicated. For instance, in the RP data used in the current study, the occupations were reported in three categories: public, private, and business (i.e., self-employed). However, depending on job type, the starting time of offices and the working hours very often vary within the single market segment. For example, in Dhaka, the opening times of public banks, administrative offices, etc. are 10:00 am, whereas public universities/schools/colleges have different start times. Therefore, it is not worthwhile to consider a constant time for a specific market segment in such a complex situation. Hence, we considered a latent desired time of travel for each market segment and estimated the parameters of the distribution of the PDT along with the other model parameters.

2.4. Theoretical model

Our modelling framework was based on the random utility framework. Random utility theory suggests that individual decision is followed by rationality and complete information. Decision-makers choose each alternative time with the highest utility, where the utility of an alternative i to a person n has the form:

$$u_n(i) = u(x_{in}, s_n, \beta) \tag{3}$$

where, x_{in} is the vector of the attribute of alternative i and for individual n, s_n is the vector of characteristics of person n and β is the parameter vector that would be estimated using the available choice data.

McFadden (1973) proposed that this utility has the linear-in-parameters separable form:

$$u(x_{in}, s_n, \beta) = V(x_{in}, s_n, \beta) + \varepsilon_{in} \tag{4}$$

where, V is the observed component of utility and ε_{in} is the unobserved error term. Based on the schedule delay theory, in our model, the generic form of the observed component was expressed linearly as a function of variables available for the departure time utility equation such as travel time, corresponding schedule delay, activity duration and other sociodemographic attributes. Therefore, the equation can be expressed as follows:

$$V_{in} = \beta_{TT} * TT_{in} + \beta_{SDE} * SDE_{in} + \beta_{SDL} * SDL_{in} + \cdots \tag{5}$$

where, TT_i is the travel time at alternative i. The early and late schedule delay can be defined as:

$$SDE_{in} = \max\,(0, PDT_n - DT_i) \tag{6}$$

and

$$SDL_{in} = \max\,(DT_i - PDT_n, 0) \tag{7}$$

where, PDT_n is the preferred departure time and DT_i is the midpoint of the departure time interval of alternative time i in terms of hours (e.g., for 7:00 to 8:00 time interval DT corresponds to 7:30). In the absence of PDT_n in the available RP data, we used statistical distributions. The simultaneous estimation of both early and late schedule delays, in conjunction with PDT_n may lead to identification and optimisation challenges. This is attributable to the mutually exclusive nature of the early and late schedule delays, necessitating to incorporate the earliness and lateness term as dummy variables. The utilisation of these dummy variables introduces complications as they are dependent upon the preferred departure time, a parameter concurrently undergoing estimation. Consequently, during the optimisation process estimating the preferred departure time, the dummy variables oscillate between 0 and 1, leading to a non-continuously differentiable function. To overcome this challenge, from a behavioural perspective, it can be assumed that the earliness and lateness disutility followed a functional form that is lower around the preferred departure time and higher when departure time spreads further away from the preferred departure time. Therefore, the schedule delay is symmetric in terms of earliness and lateness and follow a parabolic function (as this functional form give better consistency (Bwambale et al., 2019)). After adjusting the schedule delay term, the deterministic part of the utility equation can be expressed as:

$$V_{in} = \beta_{TT} * TT_{in} + \alpha(PDT_n - DT_i)^2 \tag{8}$$

where, α is the parameter to be estimated and representing the sensitivity to delay. It is expected that α will have a negative sign.

Different assumptions about the distribution of the unobserved error term ε_{in} lead to different model structures, thus offering different functional forms for the choice probabilities. In our model, ε_{in} was assumed to be independent and identically distributed across alternatives and respondents, following a Type I Extreme Value distribution (Gumbel). Therefore, the choice probability for an individual was estimated using the multinomial logit model (McFadden, 1973). The choice probabilities for each alternative i in MNL can be expressed as (for detail see (Train, K.E., 2009)):

$$P_{in}(\beta, \alpha, PDT_n) = \frac{exp(\beta_{TT} * TT_{in} + \alpha(PDT_n - DT_i)^2)}{\sum_{j \in C_n} exp\left(\beta_{TT} * TT_{jn} + \alpha(PDT_n - DT_j)^2\right)} \tag{9}$$

where, C_n is the choice set of individual n. As the PDT_n was not observed, statistical distributions were assumed to reflect the heterogeneity of PDT across the travellers. The density of PDT_n can be defined as $f(PDT|\Omega)$, here, Ω is the vector of parameters (mean and covariance) of the distribution. In this case, choice probabilities can be estimated using the following form:

$$P_n(\beta, \alpha, \Omega) = \int_{PDT} P_{in}(\beta, \alpha, PDT_n) * f(PDT|\Omega) dPDT \tag{10}$$

Since Equation 10 does not have a closed-form, a simulated log-likelihood was used using “Halton draws” from the specified distribution (normal distribution for Johnson’s distribution and uniform distribution for truncated normal distribution) to calculate the logit probabilities, which were then averaged over the number of draws. To keep the simulation variance lower in the estimated parameter and at the same time, to reduce the computation run time, we used Halton draws. The number of draws had been gradually increased starting from 50 till they were found to be stable for different starting values. The final model was estimated with 300 Halton draws (Train, K., 2000).

$$SLL = \sum_{n=1}^{N} \sum_{j=1}^{J} d_{nj} ln\, \check{P}_{nj}(\beta, \alpha, PDT_n) \tag{11}$$

where,

$$\check{P}_{in} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta, \alpha, PDT_n)^r \tag{12}$$

where, R is the number of draws and $\check{P}_{in}$ is the unbiased estimation of P_{in}. For the PDT distribution, we used Johnson’s S_B distribution for office employees and truncated normal distribution bounded between the limits of the analysis period (morning to evening) for the self-employed personnel. The S_B distribution was preferred for the former one as it has a more flexible functional form than truncated normal and log-normal distribution (Train, K. and Sonnier, 2005).

2.5. Results and discussion

Models were estimated using the "Apollo" package R, applying the Maximum Likelihood Estimation with the BFGS optimization algorithm (Hess and Palma, 2019). Model estimation was done for both outbound and return commuting trips separately. Base outbound and return models were developed first, which were simple MNL models. The base models were then extended to advanced MMNL models that acknowledged the heterogeneity in preferred departure time. Both in the base and advanced MMNL models, the effects of different socio-demographics were considered. These effects were allowed to vary among the alternative time-periods. It is important to note that in the survey data, respondents who work at the office was categorised as public and private employees and no further details about the profession were recorded. We aggregated these two categories to a single category ('office employees') as we were modelling the choices of the car and ride-hailing commuters (who are likely to be 'white collar' workers). Developed outbound and return models (both MNL and MMNL) included three broad types of independent variables: individual socio-demographics, household-level socio-demographics, and level-of-service attributes. Individual socio-demographic variables included in the model were gender, usage of ride-hailing service[12] (a proxy of car availability), observed activity duration (<3hours or not). Since in the data employees' job flexibility information was missing, we included observed activity duration as a dummy variable to capture its effect on the departure time choice. Household socio-demographics explored in the model specification included a dummy variable for income (household income >60,000BDT per month or not). The level of service attribute included travel time which was estimated for different alternative periods where the observed and unobserved travel times were calculated using the sub-model described in section 2.3.2. To distinguish the influence of different occupations, at the outbound and return MMNL model different distributions of PDT were defined. The modelling results from the outbound (home-to-work) and return (work-to-home) model are shown respectively in Table 2-3 and Table 2-4. The signs of the parameter estimates such as negative sensitivity to travel time and schedule delay were plausible, thereby aligning with the hypotheses. Furthermore, these results demonstrated consistency with findings from prior studies conducted in diverse contexts, encompassing both Global South and north regions. Nevertheless, it is important to note that the magnitude of variation in sensitivity, and its potential impact across diverse socio-demographic groups within the chosen context, revealed distinctive patterns compared to those observed in other investigated cases. Significance of each variable has been systematically discussed and validated in the subsequent sections of the study. Most of the variables considered were statistically significant at 95% confidence interval (Table 2-3 and Table 2-4). However, some parameters which were not statistically significant, corresponding coefficients were retained in the model for the sake of comparison between simple MNL and MMNL model, and intuitive interpretation of each co-efficient.

[12] Ride-hailing services include *uber* or *pathao* car use and considered as a proxy of car-availability.

2.5.1. Outbound model

For the outbound model, 10 discrete time intervals were grouped under four broad discrete time intervals such as early morning (before 7:00), morning peak (7:00 – 10:00), morning off-peak (10:00 –12:00), afternoon off-peak (12:00 – 16:00) and evening (after 16:00). These groupings were done considering the sign and magnitude of more disaggregate time-period-specific parameters.

Overall, most of the socio-demographic variables considered in the two specifications of the outbound models showed similar trends. It was observed that different socio-demographic determinants had a significant influence in determining the departure time choice of car commuters for the home-to-work trips. For example: the results from both outbound models showed that compared to male commuters or female self-employed personnel, the utility associated with departing for the outbound trip was larger for the morning peak (7:00 – 10:00) and morning off-peak (10:00 – 12:00) for female office commuters compared to early morning, afternoon off-peak and evening (Table 2-3). This result was intuitive due to the potential consequence of increased obligation of housework and safety concerns[13]. A similar trend was observed in both models that for office commuters with monthly income greater than 60,000BDT - the utility of departure for outbound was the highest during the morning peak which was followed by morning off-peak time compared to the other periods. On the contrary, for self-employed personnel with monthly income greater than 60,000BDT – the utility was the highest during the morning off-peak followed by morning peak. This implied the greater privilege of self-employed personnel to avoid peak time congestion and travel in the period of reduced travel time. In both models, office commuters and self-employed personnel who had activity durations of less than three hours had the lowest utility to travel during the morning peak time. The influence of observed activity duration can be explained by the fact that shorter activity durations (<3 hours) could insinuate the flexibility of employees both at work as well as at home. The effect of travel time was captured using generic coefficients for all time periods in both specifications. In both cases, the coefficient of travel time was negative as expected indicating disutility associated with longer travel times. The travel time variable was statistically not significant in the MNL model, but significant in the MMNL model. Therefore, it was retained in the MNL model.

The MNL and MMNL specifications however led to different sensitivities between car and ride-hailing[14] service users. In the MNL specification, the self-employed personnel using ride-hailing services showed higher utility to travel in the morning peak-time (7:00 – 10:00). While the effect of schedule delay and preferred departure times were accounted for in the MMNL specification, they showed disutility associated with travelling in the morning peak and off-peak compared to other alternative time period.

It may be noted that the influences of other socio-demographic variables (e.g., age, household size, vehicle ownership, etc.) were also tested in both specifications, but not included in the final models as their influences were not significantly different from zero. Similarly, the effects of Alternate Specific Constants (ASCs) were also tested, but not found to be significantly different from zero.

In the MMNL model, the inclusion of schedule delay led to a significant gain in model fit. The negative coefficient of schedule delay term captured the increased disutility associated with a late arrival (i.e., after the start of the office hours) or early arrival (i.e., before the start of the office). The coefficient was

[13] Majority of the cars in Dhaka are driven by male chauffeurs.
[14] The majority of the ride-hailing service users are male individuals (83%), office employees (65%), with an income below 60,000 BDT (64%) and 74% of them do not own a car.

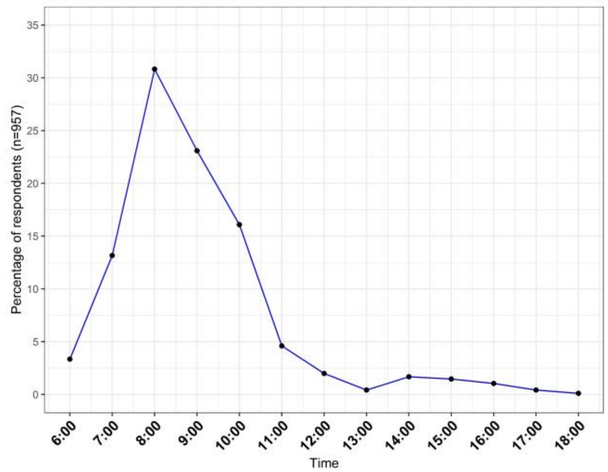
statistically significant for both occupation group, but the sensitivity to the schedule delay was slightly higher for the office employees (β = -0.09361) compared to self-employed personnel (β = -0.07999). The density curve of the preferred departure time derived from the outbound MMNL model outputs showed a single peak for office employees with approximately a mean value of 9:00 (Figure 2-2 (b)). On the other hand, for self-employed personnel, the PDT graph was almost similar to office employees (Figure 2-2 (b)) with a slight shift of mean value towards the right (around10:00). This can be explained by the fact that self-employed personnel can avoid office rush hours due to their increased flexibility and lower sensitivity to schedule delay. Further, higher sensitivity to schedule delay among office employees can be attributed to the strict enforcement of the reporting time at the workplace (i.e., requirement to sign-in upon arrival) compared to the more flexible schedule of self-employed personnel. It may be also noted that corresponding standard deviation was statistically significant only among the office employees. This can be attributed to the fact that the start times (and subsequently reporting times) of office employees working in the public and private sector are different.

Table 2-3 Base model (estimates from simple MNL model.

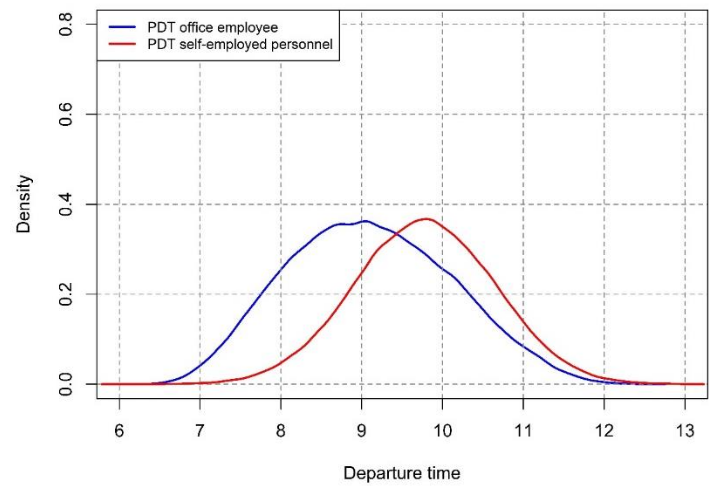
Outbound (Home-to-work) base MNL model			**Outbound (Home-to-work) advanced MMNL model**		
Parameter	Estimates	t-ratio	Parameter	Estimates	t-ratio
Socio-demographic variables					
Female-office employee			*Female-office employee*		
Morning peak (7:00 – 10:00)	1.715	4.02***	Morning peak (7:00 – 10:00)	0.981	2.63***
Morning off-peak (10:00 – 12:00)	1.811	3.98***	Morning off-peak (10:00 – 12:00)	0.956	2.15***
High income office employee			*High income office employee*		
Morning peak (7:00 – 10:00)	2.879	12.64***	Morning peak (7:00 – 10:00)	1.536	5.86***
Morning off-peak (10:00 – 12:00)	1.604	5.88***	Morning off-peak (10:00 – 12:00)	0.672	2.16***
High income self-employed personnel			*High income self-employed personnel*		
Morning peak (7:00 – 10:00)	1.873	9.34***	Morning peak (7:00 – 10:00)	0.755	3.37***
Morning off-peak (10:00 – 12:00)	2.048	9.56***	Morning off-peak (10:00 – 12:00)	1.172	4.89***
Ride-hailing service user office employee			*Ride-hailing service user office employee*		
Morning peak (7:00 – 10:00)	1.167	3.64***	Morning off-peak (10:00 – 12:00)	-0.839	-2.02***
Ride-hailing service user self-employed personnel			*Ride-hailing service user self-employed personnel*		
Morning peak (7:00 – 10:00)	0.484	1.68**	Morning peak (7:00 – 10:00)	-0.498	-1.36*
			Morning off-peak (10:00 – 12:00)	-0.698	-1.67**
Activity duration under 3 hours (office employee)			*Activity duration under 3 hours (office employee)*		
Morning peak (7:00 – 10:00)	-2.382	-3.75***	Morning peak (7:00 – 10:00)	-2.388	-4.48***
Activity duration under 3 hours (self-employed personnel)			*Activity duration under 3 hours (self-employed personnel)*		
Morning peak (7:00 – 10:00)	-0.643	-1.71**	Morning peak (7:00 – 10:00)	-0.790	-2.08***
Level of service variable					
Travel Time (minute) (β_{TT})	-0.006	-1.12	*Travel Time (minute)* (β_{TT})	-0.067	-7.15***
Situational constraint					
			Mean PDT of office employees, μ_{office}	-0.252	-2.09***
			Std. Dev. of PDT of office employees, σ_{office}	0.642	3.22***
			Schedule delay of office employees,	-0.093	-5.36***
			Mean PDT of self-employed personnel, μ_{se}	9.776	49.80***
			Std. Dev. of PDT of self-employed personnel, σ_{se}	0.872	1.45
			Schedule delay of self-employed personnel	-0.080	-4.07***

Final LL	-1888.658	LL (final)	-1739.175
Rho-square (0)	0.1429	Rho-square (0)	0.2107
Adj.Rho-square (0)	0.1379	Adj.Rho-square (0)	0.2026
AIC	3799.32	AIC	3514.35
BIC	3852.82	BIC	3601.90
Estimated parameters	11	Estimated parameters	18

*** Estimates are significant at 95% level of confidence, ** Estimates are significant at 90% level of confidence, * Estimates are significant at 80% level of confidence



(a) Departure time of home-to-work trip



(b) Preferred departure time of home-to-work trip

Figure 2-2 Observed vs preferred departure time distribution (home-to-work trip).

2.5.2. Return model

Table 2-4 summarises the estimation results of the return model. In the return model, 9 discrete time intervals were considered in the choice set: 10:00 – 12:00, 12:00 – 14:00, 14:00 – 16:00, 16:00 – 17:00, 17:00 – 18:00, 18:00 – 19:00, 19:00 – 20:00, 20:00 – 22:00 and 22:00 – 24:00. Unlike the outbound model, the Alternate Specific Constants (ASCs) of most of these time intervals were found to be significantly different from zero and hence, retained in the model.

Similar to the outbound model, the return choice set was further grouped under four broad discrete time intervals such as morning off-peak (10:00 – 12:00), afternoon off-peak (12:00 – 16:00), evening peak (16:00 – 19:00) and evening off-peak (19:00 – 22:00) to capture the heterogeneity associated with departure time choice among different socio-demographic groups. The same set of socio-demographic variables tested for the outbound trips were tested in this regard. All the socio-demographic variables considered in the two specifications of the return models (MNL and MMNL) showed similar trends. Estimated shifts in the time period parameters indicated that for both specifications, the utility for departing during afternoon off-peak (12:00 – 16:00) and evening peak (16:00 – 17:00) was higher for female commuters. The lack of a tendency among female commuters to travel in the evening could be attributed to safety concerns. Farina et al. (2022) elucidated that female travellers tend to limit their nighttime travel due to perceived safety issues, necessitating extensive planning, preparation, and reliance on both financial and non-financial resources.

In terms of the availability of a personal car for return trips, the differences in preferences were tested separately for office workers and self-employed people. The shifts in the time period parameter estimates were statistically significantly different between the office employee and self-employed personnel – possibly due to the higher rate of car ownership (and hence lower propensity to use ride-hailing services) among the self-employed personnel. The estimates revealed that office employees had higher utility to travel during evening off-peak time (19:00 – 22:00) for return trips if they were using car-based ride-hailing services compared to the self-employed personnel and office employee using personal car. This was likely to be driven by the propensity to avoid the peak surcharge.

In terms of income, from both MNL and MMNL model specifications, it was found that for the office employee group with monthly income greater than 60,000BDT, the utility of returning was the highest during the evening peak followed by the afternoon off-peak and evening off-peak. On the other hand, for the self-employed personnel, those who had monthly income greater than 60,000BDT, the utility was the highest for departing at afternoon off-peak followed by evening off-peak and evening peak compared to other alternatives. This was likely to be associated with the higher flexibility of schedule of the self-employed group. For office commuters and self-employed personnel who had observed activity duration of less than three hours, the utility for travelling during the peak time (16:00 – 19:00) was the lowest. The disutility of longer travel time of peak period might have exceeded the utility associated with short duration activity participation.

Unlike the outbound model, travel time co-efficient was a significant determinant both in the return trip MNL and MMNL model. However, it was evident that the travel time parameter had a greater influence on the outbound trip ($\beta = -0.06737$ in the MMNL model) compared to the return trip ($\beta = -0.009347$ in the MMNL model). This suggested that car commuters were less willing to spend longer time in traffic for outbound trips as compared to return trips.

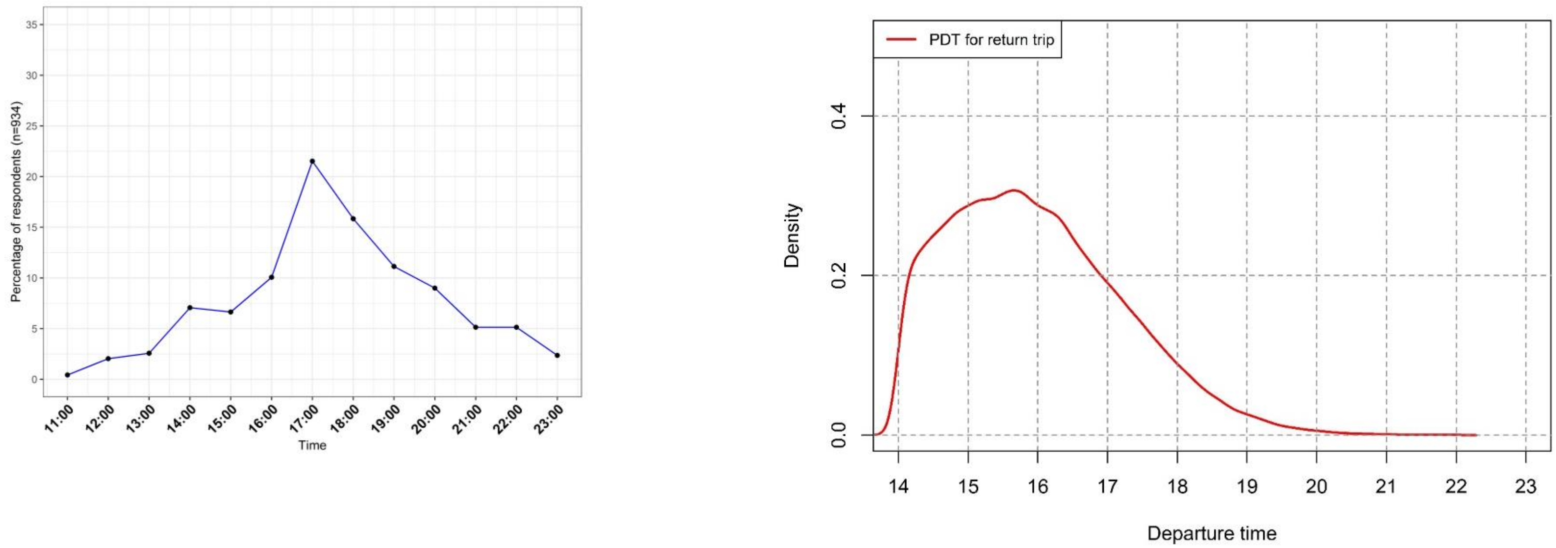
Unlike the outbound MMNL model, the inclusion of situational constraint on the return MMNL model did not lead to a statistically significant improvement in the model fit and the coefficient of schedule delay was not statistically different from zero. This can be attributed to the fact that there was more flexibility in the schedule during the return segment. The parameters of the preferred departure time distribution (mean and standard deviation) were not found to be statistically significant reflecting this flexibility. Though in the return segment MNL model outperformed the MMNL model, the density curve of the preferred departure time derived from the return MMNL model was consistent with the reality (Figure 2-3 (b)). It is also observed that the sensitivity to schedule delay was generally higher during the outbound when compared to the return trip. This was intuitive because late arrival at the office probably had more serious consequences or penalties than late arrival at home. Overall, the model fit of the return model had a lower R-square value compared to the outbound model which might be because of a large number of alternatives (times) considered in the model specification. Since the inclusion of situational constraints on the return MMNL model did not improve the model performance, this study recommended the MNL model for the return segment.

Table 2-4 Final model (estimates from mixed logit model).

Return (Work-to-home) base MNL model			**Return (Work-to-home) advanced MMNL model**		
Baseline time period constants	Estimates	t-ratio	Baseline time period constants	Estimates	t-ratio
ASC (10:00 – 12:00)	0	n/a	ASC (10:00 – 12:00)	0	n/a
ASC (12:00 – 14:00)	0.540	1.58*	ASC (12:00 – 14:00)	0.529	1.42
ASC (14:00 – 16:00)	1.593	5.11***	ASC (14:00 – 16:00)	1.577	5.26***
ASC (16:00 – 17:00)	1.491	4.96***	ASC (16:00 – 17:00)	1.475	5.06***
ASC (17:00 – 18:00)	2.302	7.89***	ASC (17:00 – 18:00)	2.287	8.21***
ASC (18:00 – 19:00)	2.006	6.79***	ASC (18:00 – 19:00)	1.993	7.08***
ASC (19:00 – 20:00)	1.790	5.90***	ASC (19:00 – 20:00)	1.782	6.06***
ASC (20:00 – 22:00)	2.033	6.61***	ASC (20:00 – 22:00)	2.034	6.68***
ASC (22:00 – 24:00)	1.662	5.65***	ASC (22:00 – 24:00)	1.679	4.94***
Shifts in time period constants			*Shifts in time period constants*		
Female-office employee			*Female-office employee*		
Afternoon off-peak (12:00 – 16:00)	2.195	2.94***	Afternoon off-peak (12:00 – 16:00)	2.194	2.65***
Evening peak (16:00 – 19:00)	1.605	2.20***	Evening peak (16:00 – 19:00)	1.604	2.03***
Evening off-peak (19:00 – 22:00)	1.220	1.62*	Evening off-peak (19:00 – 22:00)	1.220	1.50*
Female-self-employed personnel			*Female-self-employed personnel*		
Afternoon off-peak (12:00 – 16:00)	1.992	1.88**	Afternoon off-peak (12:00 – 16:00)	1.991	2.22***
Evening peak (16:00 – 19:00)	1.780	1.67**	Evening peak (16:00 – 19:00)	1.780	1.92**
Evening off-peak (19:00 – 22:00)	1.129	1.01	Evening off-peak (19:00 – 22:00)	1.128	1.19
High income office employee			*High income office employee*		
Afternoon off-peak (12:00 – 16:00)	0.555	1.61**	Afternoon off-peak (12:00 – 16:00)	0.556	1.99***
Evening peak (16:00 – 19:00)	0.979	3.22***	Evening peak (16:00 – 19:00)	0.980	3.71***
Evening off-peak (19:00 – 22:00)	0.306	0.93	Evening off-peak (19:00 – 22:00)	0.303	1.08
High income self-employed personnel			*High income self-employed personnel*		
Afternoon off-peak (12:00 – 16:00)	0.711	2.17***	Afternoon off-peak (12:00 – 16:00)	0.711	2.17**
Evening peak (16:00 – 19:00)	0.260	0.86	Evening peak (16:00 – 19:00)	0.260	0.65
Evening off-peak (19:00 – 22:00)	0.680	2.14***	Evening off-peak (19:00 – 22:00)	0.680	1.78**
Ride-hailing service user office employee			*Ride-hailing service user office employee*		
Evening off-peak (19:00 – 22:00)	0.795	2.97***	Evening off-peak (19:00 – 22:00)	0.795	3.03***
Ride-hailing service user self-employed personnel			*Ride-hailing service user self-employed personnel*		
Evening peak (16:00 – 19:00)	-0.656	-1.89**	Evening peak (16:00 – 19:00)	-0.656	-1.85**

Evening off-peak (19:00 – 22:00)	-0.531	-1.37*	Evening off-peak (19:00 – 22:00)	-0.531	-1.38*
Activity duration under 3 hours (office employee)			*Activity duration 3 hours (office employee)*		
Evening peak (16:00 – 19:00)	-1.742	-3.11***	Evening peak (16:00 – 19:00)	-1.742	-3.07***
Evening off-peak (19:00 – 22:00)	-1.390	-2.03***	Evening off-peak (19:00 – 22:00)	-1.390	-1.99***
Return (Work-to-home) base MNL model			**Return (Work-to-home) advanced MMNL model**		
Parameter	Estimates	t-ratio	Parameter	Estimates	t-ratio
Activity duration under 3 hours (self-employed personnel)			*Activity duration under 3 hours (self-employed personnel)*		
Evening peak (16:00 – 19:00)	-2.195	-3.93***	Evening peak (16:00 – 19:00)	-2.194	-3.89***
Evening off-peak (19:00 – 22:00)	-2.190	-3.47***	Evening off-peak (19:00 – 22:00)	-2.190	-3.46***
Level of service variable					
Travel Time (minute) (β_{TT})	-0.009	-2.11***	*Travel Time (minute)* (β_{TT})	-0.009	-2.06***
Situational constraint					
			Preferred Departure time (μ)	15.487	0.44
			Preferred Departure time (σ)	1.577	0.25
			Schedule delay	-0.0006	-0.67
Final LL		-1848.716	LL (final)		-1848.716
Rho-square (0)		0.0992	Rho-square (0)		0.0992
Adj.Rho-square (0)		0.0855	Adj.Rho-square (0)		0.0841
AIC		3753.43	AIC		3759.43
BIC		3888.94	BIC		3909.45
Estimated parameters		28	Estimated parameters		31

*** Estimates are significant at 95% level of confidence, ** Estimates are significant at 90% level of confidence, * Estimates are significant at 80% level of confidence



(a) Departure time of work-to-home trip (b) Preferred departure time of work-to-home trip

Figure 2-3 Observed vs preferred departure time distribution (work-to-home trip).

2.5.3. Time value of schedule delay

Generally, commuters encounter scheduled disutility due to early arrival or late arrival at the workplace. Hence, commuters attempt to choose the appropriate departure time by making a trade-off between travel time and schedule delay. For instance, they can choose the alternative with the 'best travel time but large schedule delay' or 'worst travel time with no schedule delay' or anything in between. Hence, to get better insight from our developed model, we further estimated the time valuation of schedule delay (TVSD) to understand the sensitivities to schedule delay versus travel time. TVSD was estimated using the formulation proposed by Bwambale et al. (2019). It was estimated as the ratio of partial derivatives of the utility equation (equation 8) with respect to schedule delay and travel time. The following equation (13) was used, and the average of the estimated output is shown in Table 2-5. Since the schedule delay term was insignificant for the return trip, TVSD was only estimated for the outbound trip.

$$TVSD_{occupation} = \frac{\partial V_{in}/\partial SD_{in}}{\partial V_{in}/\partial TT_{in}} = \frac{\alpha_{occupation} * 2(PDT_n - DT_i)}{\beta_{TT}} \qquad (13)$$

The estimated TVSD is the unitless metric that represents the amount of delay a commuter was willing to experience for a unit reduction in travel time. The TVSD was lower for office employees compared to self-employed personnel. This value signified the fact that office employees had very less willingness to accept schedule delays – potentially due to the very inflexible working hours in the office. On the other hand, the self-employed personnel had the flexibility to choose longer schedule delays for the sake of reducing travel time. It may be noted that it was not possible to validate these due to the absence of a supplementary source of information. However, according to the literature, the time valuation of schedule delay in European countries varies between 0.81 to 1.71 (Wardman et al., 2012). Hence, the estimated output seems logical as the time valuation in developing countries will be lower than the developed countries.

Table 2-5 Time valuation of schedule delay.

Direction	Office employee	Self-employed personnel	Weighted Average
Outbound	0.103	0.83	0.422

2.6. Policy insights

The estimated model parameters can be utilised in formulating peak spreading policies for car travel in Dhaka, Bangladesh. To utilise the model for comprehending the sensitivity of various market segments for policy formulation, we conducted arbitrary forecasting analyses. In each scenario, every demographic group was permitted to dominate the choice scenario. For instance, in one scenario, all respondents were assumed to behave like office employees. Similarly, other demographics considered in the model development, such as female commuters, Uber usage, reduced working hours, and household income exceeding 60,000 BDT, were modified in the forecasting analysis. Some proposals have been suggested below based on the parameter estimates and forecasting analysis:

- The coefficients of gender revealed that female office employees had a higher utility to travel by car during the morning and evening peak hours. The forecasting analysis also indicates that if all travellers are female commuters, there will be a shift from morning outbound off-peak alternatives to peak periods (8:00 – 9:00 and 9:00 – 10:00) (Table 2-6). Conversely, for the return journey, a shift from the evening to the afternoon (before 18:00) can be observed (Table 2-7).

Therefore, targeted incentives for female employees (e.g., flexibility of the start times, working from home privileges, discounted transport cost during the off-peak and/or public transport, etc.) and safety measures for evening travel can hence be taken under consideration.

- The coefficients of income revealed that high-income office employees (household income >60,000BDT per month) had the highest utility to travel for work trips during the peak time when the schedule delay was minimal, but travel time was the worst. According to the forecasting, it is evident that for outbound commuters, there will be a 28% increase in travel demand between 7:00 to 10:00 if all travellers are from high-income households. Similarly, for the return journey, a significant rise can be observed between 17:00 to 19:00 compared to other alternative return periods. This demographic group had higher affordability and hence was less likely to be price sensitive. Therefore, a congestion pricing policy targeting the peak time traveller, though an effective way for revenue generation, may not be effective in shifting this group from travelling in the peak. This revenue can be a useful means of funding efficient and dependable public transport services.
- The utility of traveling during peak and off-peak time was very subtle for the users of ride-hailing services (e.g., Uber, Pathao Car). For instance, the forecasting indicates a heightened tendency to travel during the morning peak if all travellers utilise ride hailing services, whereas, for the return journey, the increase in demand occurs during the evening off-peak period. Therefore, changing the pricing structure of ride-hailing services to make peak travel much more expensive compared to off-peak can serve as an incentive to travel during the off-peak time.
- The higher sensitivity to schedule delay of the office employee and their lower willingness to accept schedule delay (i.e., lower time value of schedule delay) reflected the strictness of their schedule. In outbound forecasting, if all travellers are office employees, a significant shift in travel demand can be observed before 10:00, which aligns with the typical office start time in Bangladesh. However, for the return journey, such a shift towards the office closing time at 17:00 was not as pronounced. This also highlights the greater significance of schedule delays in the morning compared to the return journey. Hence, enhancing the flexibility of working hours of office employees (e.g., staggered start and end times, flexible start times, etc.) is a critical pre-requisite before the implementation of congestion pricing policies. To do so, the transport authority can work in collaboration with the employers to estimate the required ratio of employees needed at a time in the office premises and offer flexibility to the rest so that they can travel during the off-peak time if needed.
- The schedule delay was not found to have a significant effect on the work-to-home trips. Therefore, the afternoon peak is likely to be easier to flatten compared to the morning peak.

Table 2-6 Forecasting scenario results (Outbound).

	Sample	Base	All travellers belong to high-income group	All travellers are office employee	All travellers use uber	Work duration of all travellers is less than 3 hours	All travellers are female
6:00 - 7:00	5.12	0.06	0.04	0.07	0.08	0.13	0.04
7:00 - 8:00	13.17	0.19	0.2	0.22	0.19	0.09	0.2
8:00 - 9:00	30.83	0.23	0.26	0.27	0.25	0.11	0.25
9:00 - 10:00	23.09	0.19	0.21	0.21	0.2	0.09	0.22

10:00 - 11:00	16.09	0.16	0.16	0.12	0.11	0.3	0.16
11:00 - 12:00	4.6	0.08	0.08	0.05	0.05	0.13	0.06
12:00 - 14:00	2.4	0.06	0.04	0.05	0.08	0.11	0.05
14:00 - 16:00	3.13	0.02	0.01	0.01	0.03	0.03	0.02
16:00 - 17:00	1.04	0.01	0	0	0.01	0.01	0
17:00 - 18:00	0.52	0	0	0	0	0	0

Table 2-7 Forecasting scenario results (Return).

	Sample	**Base**	**All travellers belong to high-income group**	**All travellers are office employee**	**All travellers use uber**	**Work duration of all travellers is less than 3 hours**	**All travellers are female**
10:00 - 12:00	1.5	0.01	0.01	0.01	0.02	0.04	0
12:00 - 14:00	4.6	0.05	0.05	0.04	0.05	0.11	0.07
14:00 -16:00	13.7	0.14	0.14	0.12	0.15	0.34	0.21
16:00-17:00	10.06	0.1	0.1	0.11	0.08	0.04	0.11
17:00-18:00	21.41	0.21	0.22	0.24	0.17	0.09	0.24
18:00-19:00	15.85	0.16	0.17	0.18	0.13	0.07	0.17
19:00-20:00	11.13	0.11	0.11	0.1	0.14	0.06	0.08
20:00-22:00	14.13	0.14	0.14	0.13	0.18	0.07	0.1
22:00-24:00	7.6	0.08	0.06	0.07	0.08	0.18	0.02

2.7. Conclusion

In this research, the key challenges in modelling the departure time choice model in the context of Dhaka, Bangladesh were identified, and solutions were proposed. Separate departure time choice models of home-to-work and work-to-home of commuters using personal car/ride-hailing services were developed to demonstrate the proposed solution approached to overcome the limitations of using RP data in modelling departure time choice. The methodological contributions include the following:

(1) A new method to estimate the travel time for the full range of alternative time periods using Google Maps API and stated travel times when the Google Maps API is not deemed to be a reliable stand-alone source of travel time information.

(2) Extension of the state-of-the-art method for representing PDT. Instead of assuming a constant value for a specific market segment or a generic statistical distribution, the proposed method included two different statistical distributions for office workers and self-employed people acknowledging the high level of heterogeneity between and within each group. The estimation results supported the hypothesis that a significant difference exists among different occupation groups in terms of their departure time choice.

Based on the results, an advanced MMNL model was recommended for outbound trips to account for the heterogeneity in schedule delay among the travellers. A simple MNL model was found to be adequate for the return trip segment where the schedule delay was not found to have a significant effect. The key aspects of the study are listed below:

- The estimation results provided empirical evidence that departure time choices in Dhaka were significantly affected by activity duration, and schedule delay in addition to travel time. The results also revealed substantial heterogeneity depending on the type of job.
- The results indicated that preferred departure times/arrival times, though unobserved in the RP data, were important aspects of departure time choice models. The proposed modelling framework to estimate the unobserved preferred departure time through the assumed distribution parameters (mean and standard deviation) using a mixed logit framework can be an effective way to address the unobserved preferred departure time issue, even in cross-sectional data. The framework can be also applied in the case of passively generated data sources (e.g., GPS, mobile phone data, etc.) as well which also has the unobserved preferred departure time problem.
- Along with the distribution parameters, we estimated the sensitivities to schedule delay of different occupation groups which can be critical inputs in designing effective peak spreading policies in Dhaka city. Results highlighted the fact that schedule delay and preferred departure time parameters were significant in the home-to-work trip, but not in the work-to-home trip segments. This finding can have important policy implications.
- Further, the result suggested that car commuters were sensitive to travel time for both outbound and return trips. Therefore, policies aiming to reduce traffic congestion such as road pricing, inbound flow control, etc. will enable commuters to adopt their preferred departure time at the expense of minimal schedule delay.
- Results indicated that schedule delay was the dominant factor for home-to-work trips and the time-value of schedule delay was much less compared to European countries (i.e., a commuter is willing to accept less unit of schedule delay per unit reduction in travel time). The effect of income (high-income office employee dummy) was found to be more substantial than that of gender (female office employee dummy) for the home-to-work trips – but the trend was opposite in the case of the return trip. The results thus reinforced the notion regarding the problems associated with the transferability of the models between developing countries.

It may be noted that the effect of travel cost was explored as well. But travel cost was not recorded in the data and in the absence of information about the vehicle type and the type of the driver[15] it was not possible to estimate the cost in a reliable manner. This can be explored in the future using primary data or appropriate supplementary data. Further, the current study focused only on commute trips made by car which are the biggest contributors to traffic congestion in Dhaka. In the future, this can be extended to commute trips made by public transport and paratransit modes and include other trip purposes.

However, even in its current form, the research findings can be practically useful for devising peak-spreading policies in Dhaka – either as a stand-alone tool to test the impact of varied start-times of offices in different locations or within an agent-based simulation tool to test the impact of different congestion pricing policies. In addition, the proposed framework can be useful in other developing countries with similar data issues.

[15] Most cars in Dhaka are chauffeur-driven and there can be a substantial variation in the cost depending on the skill-level of the chauffer.

2.8. References

Abedin, N., Rahman, M.M., Hossain, M.I., Hisazumi, K. and Ahmed, A. 2020. Travel behavior of SME employees in their work commute in emerging cities: A case study in Dhaka City, Bangladesh. *Sustainability.* **12**(24), p10337.

Anwar, A.M. 2012. Departure time choice analysis using ordered response probit model: journey to work trip in Dhaka City of Bangladesh.

Arellana, J., Daly, A., Hess, S., de Dios Ortúzar, J. and Rizzi, L.I. 2012. Development of surveys for study of departure time choice: two-stage approach to efficient design. *Transportation Research Record.* **2303**(1), pp.9-18.

Arellana, J., de Dios Ortúzar, J. and Rizzi, L.I. 2013. Survey data to model time-of-day choice: methodology and findings. *Transport Survey Methods: Best Practice for Decision Making.* Emerald Group Publishing Limited.

Bangladesh, B.P.i. 2023. [Online]. Available from: https://www.lloydsbanktrade.com/en/market-potential/bangladesh/opening-hours

Ben-Akiva, M. and Abou-Zeid, M. 2013. Methodological issues in modelling time-of-travel preferences. *Transportmetrica A: Transport Science.* **9**(9), pp.846-859.

Bhat, C.R. 1998. Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological.* **32**(6), pp.361-371.

Bhat, C.R. and Steed, J.L. 2002. A continuous-time model of departure time choice for urban shopping trips. *Transportation Research Part B: Methodological.* **36**(3), pp.207-224.

Bwambale, A., Choudhury, C.F. and Hess, S. 2019. Modelling departure time choice using mobile phone data. *Transportation research part A: policy and practice.* **130**, pp.424-439.

Cervero, R. 2013. Linking urban transport and land use in developing countries. *Journal of transport and land use.* **6**(1), pp.7-24.

Chaichannawatik, B., Kanitpong, K. and Limanond, T. 2019. Departure time choice (DTC) behavior for intercity travel during a long-holiday in Bangkok, Thailand. *Journal of Advanced Transportation.* **2019**.

De Jong, G., Daly, A., Pieters, M., Vellay, C., Bradley, M. and Hofman, F. 2003. A model for time of day and mode choice using error components logit. *Transportation Research Part E: Logistics and Transportation Review.* **39**(3), pp.245-268.

Dimitriou, H.T. and Gakenheimer, R. 2011. Urban transport in the developing world: A handbook of policy and practice.

Dong, H. and Cirillo, C. 2020. Space-time dynamics: A modeling approach for commuting departure time on linked datasets. *Journal of Transport Geography.* **82**, p102548.

DTCA. 2015. *Revised strategic transport plan (RSTP), Dhaka, Bangladesh.* Dhaka: Dhaka Transport Coordination Authority (DTCA).

DTCB. 2005. *Strategic Transport Plan (STP) for Dhaka, Final Report.* Dhaka Transport Coordination Board (DTCB), Dhaka, Bangladesh.

Enam, A. and Choudhury, C.F. 2011. Methodological issues in developing mode choice models for Dhaka, Bangladesh. *Transportation research record.* **2239**(1), pp.84-92.

Fantom, N.J. and Serajuddin, U. 2016. The World Bank's classification of countries by income. *World Bank Policy Research Working Paper.* (7528).

Hashi, E.K., Hasan, M.R. and Zaman, M.S.U. 2015. A heuristic solution of the Vehicle Routing Problem to optimize the office bus routing and scheduling using Clarke & Wright's savings algorithm. In: *2015 international conference on computer and information engineering (ICCIE)*: IEEE, pp.13-16.

Hendrickson, C. and Plank, E. 1984. The flexibility of departure times for work trips. *Transportation Research Part A: General.* **18**(1), pp.25-36.

Hess, S. and Palma, D. 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of choice modelling.* **32**, p100170.

Hess, S., Polak, J.W., Daly, A. and Hyman, G. 2007. Flexible substitution patterns in models of mode and time of day choice: new evidence from the UK and the Netherlands. *Transportation.* **34**(2), pp.213-238.

Holyoak, N. 2008. *Departure time choice for the car-based commute*. thesis, Planning and transport research centre (PATREC).

JICA. 2010. *DHUTS Dhaka Urban Transport Network Development Study.*

Koppelman, F.S., Coldren, G.M. and Parker, R.A. 2008. Schedule delay impacts on air-travel itinerary demand. *Transportation Research Part B: Methodological.* **42**(3), pp.263-273.

Li, H., Li, X., Xu, X., Liu, J. and Ran, B. 2018. Modeling departure time choice of metro passengers with a smart corrected mixed logit model-a case study in Beijing. *Transport Policy.* **69**, pp.106-121.

McCafferty, D. and Hall, F.L. 1982. The use of multinomial logit analysis to model the choice of time to travel. *Economic Geography.* **58**(3), pp.236-246.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior.

Monolina, P. and Ekram, K.M.M. nd. ANALYSIS OF SCHOOL TRAFFIC IN DHANMONDI RESIDENTIAL AREA.

RAJUK. 2021. *Dhaka Structure Plan 2016-2035.* Dhaka: Rajdhani Unnayan Kartripakkha (RAJUK; the Capital Development Authority).

Siddique, M.A.B., Choudhury, C.F. and Mondal, P.K. 2014. *Modelling the Behavioural Response to Congestion Pricing in Dhaka.*

Small, K.A. 1982. The scheduling of consumer activities: work trips. *The American Economic Review.* **72**(3), pp.467-479.

Thorhauge, M., Cherchi, E., Walker, J.L. and Rich, J. 2019. The role of intention as mediator between latent effects and behavior: application of a hybrid choice model to study departure time choices. *Transportation.* **46**(4), pp.1421-1445.

Train, K. 2000. Halton sequences for mixed logit.

Train, K. and Sonnier, G. 2005. Mixed logit with bounded distributions of correlated partworths. *Applications of simulation methods in environmental and resource economics.* Springer, pp.117-134.

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

Wardman, M., Chintakayala, P., de Jong, G. and Ferrer, D. 2012. European wide meta-analysis of values of travel time. *ITS, University of Leeds, Paper prepared for EIB.*

WB. 2015. *Open Transport Development and Integration of ICT and Transport.* World Bank.

Yang, F., Jin, X. and Liu, R. 2008. Tour-Based Time-of-Day Choices for Weekend Nonwork Activities. *Transportation research record.* **2054**(1), pp.37-45.

Zannat, K.E., Ashraful Islam, K., Sunny, D.S., Moury, T., Tuli, R.D., Dewan, A. and Adnan, M.S.G. 2021. Nonmotorized commuting behavior of middle-income working adults in a developing country. *Journal of Urban Planning and Development.* **147**(2), p05021011.

Zou, M., Li, M., Lin, X., Xiong, C., Mao, C., Wan, C., Zhang, K. and Yu, J. 2016. An agent-based choice model for travel mode and departure time and its case study in Beijing. *Transportation Research Part C: Emerging Technologies.* **64**, pp.133-147.

Chapter 3 Modelling time-of-travel preferences capturing correlations between departure times and activity durations

Abstract

Departure time choice models quantify the relative impacts of the factors affecting travellers' departure time selection and help design targeted peak-spreading policies. The departure time preference of travellers is traditionally captured using parameters associated with different alternatives along three aspects – outbound, return, and duration. In reality, departure time decisions for outbound and return legs, and the corresponding activity durations, are interrelated in most cases. However, none of the previous departure time choice models has explicitly investigated the impact of this potential correlation on model outputs. To address this gap in the existing literature, we proposed a model structure with a novel polynomial functional form of alternative specific constants (ASCs) that captures this correlation in a joint (outbound and return) departure time choice model. A revealed preference (RP) dataset from Dhaka, Bangladesh, was used to model joint departure time preferences of the car commuters. The proposed model was then compared with a state-of-the-art model that used a trigonometric formulation of the ASCs. Results indicated that the proposed formulation yielded more behaviourally realistic outputs compared to the trigonometric model by explicitly capturing the correlation between departure time and duration. The framework can thus help predict the departure times better and improve the formulations of the peak spreading policies. While the specific outputs are applicable to car commuters residing in Dhaka, Bangladesh, the framework can be applied to better predict departure times and improve the formulations of the peak spreading policies in other contexts as well.

Keywords: Departure time choice; Discrete Choice Model; Alternative Specific Constants (ASCs); Polynomial formulation, Dhaka; Bangladesh

3.1. Introduction

Traffic congestion in dense urban areas primarily stems from the concentration of travel demand over the peak hours. This adversely affects the quality of urban life in various ways, including reduced travel speeds, greater variability in travel times, increased uncertainty in arrival times, higher operating costs (fuel consumption), heightened levels of air and noise pollution, and decreased safety (Newbery, 2005; Li and Hensher, 2012; Thorhauge, Mikkel et al., 2016). This challenge is further exacerbated by urban population growth leading to increased travel demand, a heavy reliance on cars for mobility, and inefficient demand management strategies (Hensher and Puckett, 2007; Pucher et al., 2007; Batur and Koç, 2017). Addressing the ever-increasing demand and mitigating the negative impacts of congestion necessitates a multifaceted approach, considering both supply and demand sides. Simply expanding infrastructure is not a sustainable solution, as it often leads to induced traffic due to increased capacity (Noland and Lem, 2002; Thorhauge, Mikkel et al., 2020). As a result, urban planners worldwide are placing growing emphasis on demand-side strategies to shift transportation preferences (Guo et al., 2021; Geng et al., 2023). These strategies range from direct measures like congestion pricing (which can lead to changes in mode, departure time and destination (Saleh and Farrell, 2005; Börjesson and Kristoffersson, 2018; Li, W. et al., 2018)) to less restrictive ones like implementing flexible work hours (van der Loop et al., 2019; Munch and Proulhac, 2023), time-variant fares (Hightower et al., 2022), providing incentives to travel during off-peak hours and under-utilised routes (Pan et al., 2016), promoting mixed land use to alter activity locations and channel traffic away from the downtown, etc. (Cervero, 1991). However, despite their effectiveness, the design of policies to shift departure time has received less attention compared to strategies targeting mode shift and route choice (Hendrickson and Plank, 1984; Arellana et al., 2013; Azhdar and Nazemi, 2020; Zhou et al., 2020; Huan et al., 2021; Thorhauge, M. et al., 2021), particularly in cities from developing country (Zannat et al., 2022). This is often attributed to the challenges of quantifying the relative impacts of factors influencing travellers' temporal demand, primarily due to a lack of dependable data sources and the synergies among various activity-travel related choice dimensions (Graham et al., 2020).

Departure time choice models, mathematically or statistically formulated to define time-of-day choices as a function of trip and level of service attributes and socio-demographic characteristics, play a crucial role in determining the temporal distribution of urban transportation demand (Bhat and Steed, 2002; Habib, K.N., 2021). Researchers attempted to develop various departure time choice models influenced by Vickrey (1969). These models and their modelling frameworks vary based on their context of application. Some utilised continuous time choice models using hazard-based duration frameworks, like Bhat and Steed (2002) and Wang (1996). Others have adopted a discrete choice framework, dividing the continuous departure time variable into a finite set of discrete intervals and modelling utility as a function of level of service attributes, socio-demographic factors, and activity-related variables (Ding et al., 2015; Anowar et al., 2019; Bwambale et al., 2019; Chaichannawatik et al., 2019; Golshani et al., 2019). Small (1982), McCafferty and Hall (1982) and Holyoak (2008) used multinomial logit (MNL) model to understand commuters' departure time choices. The MNL model has also been used to predict the time-of-day choice to explore the differences between the weekday and weekend or holiday travel patterns (Yang et al., 2008; Chaichannawatik et al., 2019). Approaches like nested logit (NL), cross-nested logit (CNL), continuous CNL, and mixed multinomial logit (MMNL) relaxed the independence of irrelevant alternatives (IIA) assumption of the MNL model to accommodate the correlation between adjacent time intervals (Chin, 1990; Ben-Akiva and Bierlaire, 1999; Börjesson, 2008; Lemp et al., 2010). Application of joint choice modelling, as estimated by Hendrickson and Plank (1984), and Hossain et al. (2021) simultaneously developed time-of-day choices alongside other travel decisions, such as mode choice. Habib, K.M.N. et al. (2009) and Bhat (1998) used joint multinomial logit (MNL) and generalised extreme value (GEV) formulations for modelling mode and departure time choice models focusing on commuter and non-commuters' trips, respectively. Moreover, Li, H. et al. (2018), De Jong et al. (2003), and Hess et al. (2007) used a mixed multinomial logit (MMNL) model to investigate joint mode and departure time choices capturing the correlation between alternatives which are close to each other. Heterogeneity in time-of-day choice by different market segments is also captured by the latent class choice models (Thorhauge, M. et al., 2021). Additionally, Bayesian network and machine learning models have been explored for time-of-day choice analysis (Zhu et al., 2018).

The discrete choice models of departure time involve trade-offs between the time of day and associated travel time and costs. Outside of peak hours, the travel times are shorter, the congestion levels are lower, and the travel costs are often lower (e.g., off-peak public transport tickets). However, there can be an indirect cost associated with less convenient departure times, captured as "Schedule Delay" (Börjesson, 2008; Börjesson, 2009). Due to the difficulties in simulating the preferred departure times, the schedule delay-based technique performs well for exploratory modelling but is challenging for long-term forecasting applications (Hess et al., 2005). For the forecasting application, a more straightforward approach uses constants associated with different time periods to represent travellers' time preferences. However, specifying these constants is complex due to the number and length of time periods considered, limiting its applicability (Ben-Akiva and Abou-Zeid, 2013). Previous studies have used either a small number of coarse time periods or a large number of fine time periods, leading to increased computational costs and parameter identification issues. To address such issues, different studies have proposed functional approximation of the alternative specific constants (ASC) (Hess et al., 2005; Ben-Akiva and Abou-Zeid, 2013). This approach offers several benefits to the model such as 1) reducing the computational cost by lowering the number of parameters to be estimated, 2) avoiding the identification issues associated with the discontinuities in the utility function and the absence of observations for some arrival and departure time periods in the data, and 3) facilitating the interpretation of the results. Various functional forms, such as trigonometric, piecewise linear, and power series expansion functions, have been proposed to estimate the distribution of these constants (Hess et al., 2005; Abou-Zeid et al., 2006; Ben-Akiva and Abou-Zeid, 2013). The constants used in the previous studies to capture the time preferences of travellers were related to three dimensions — outbound, return, and duration. However, their proposed specification employs two separate functional forms for outbound and return (or duration) times, thus overlooking the interdependency and correlations among departure time, duration, and return time. Inappropriate assumptions regarding the functional form may result in specification errors and introduce uncertainties in the model predictions (De Jong et al., 2007).

To accurately model the time of the day preference, consideration of correlations and interactions between departure time and duration is crucial. Positive associations between departure time and duration may suggest a preference for later time-of-day choices as duration requirements increase, while negative associations may indicate an inclination towards earlier time-of-day choices. This hypothesis is grounded in the understanding that varying schedule constraints and the flexibility of working hours influence the perceived utility of different departure times for comparable activity durations (Ashiru et al., 2004; Badiola et al., 2019). Testing this hypothesis will provide valuable insights into the nuanced dynamics of time-of-day preferences in urban travel behaviour. This is because the utility of departing at 9 am for an 8-hour work period is not expected to be the same as departing at 3 pm for the same activity duration, owing to potential schedule constraints and the flexibility of working hours. Additionally, peak-hour outbound travel demand for 8-10 hours of activity duration may lead to increased demand for return travel during peak hours. Similarly, the time-of-day choice during peak hours for a specific activity duration will not have the same impact on the network as it would during off-peak hours.

Therefore, in this study we proposed a polynomial functional form of ASCs that captures the correlation among the constants of the outbound, return (or duration) with an aim to improve the behavioural realism of the departure time choice models. The proposed structure is calibrated with data from Dhaka, one of the fastest growing megacities in the world and the capital of Bangladesh. The results of the proposed model are compared with those derived from the state-of-the-art method for capturing time of day preferences (based on the trigonometric formulation proposed by Ben-Akiva and Abou-Zeid (2013). This article makes two key contributions. Firstly, it introduces a flexible and efficient functional form of alternative specific constants (ASCs) that captures the interaction between departure time and duration preferences (which hasn't done before) exemplified with, but not limited to commuting trips using a discrete choice framework. The proposed model framework is expected to serve as an improved tool for planners and policy makers in better understanding the preferences of the travellers and designing effective peak-spreading policies to reduce peak hour travel demand or promote off-peak travel. Secondly, it focuses on Dhaka (the sixth largest megacity in the world in terms of population) as a case study and suggests specific planning and policy measures for to reduce peak hour car travel demand. Further, while the specific findings may not be transferable to other developing countries, the modelling approach will offer valuable insights to transport planners and policymakers in overcoming the challenges of developing robust departure time choice models amidst data scarcity and limited resources.

The rest of the paper is organised as follows: the following section describes the data sources used in this study. The modelling issues are presented next, followed by the description of the model structure and the estimation results. The findings and forecasting analysis are summarised in the end, along with directions for future research.

3.2. Data

The review of the literature reveals that most of the previous departure time choice models have used stated preference (SP) (De Jong et al., 2003; Hess et al., 2007; Arellana et al., 2012; Arellana et al., 2013; Thorhauge, Mikkel et al., 2019; Azhdar and Nazemi, 2020) with a lower number using revealed preference (RP) datasets (Bhat, 1998; Yang et al., 2008; Chaichannawatik et al., 2019). Even though it is easier to specify the choice set and the preferred departure time in the SP, such data may be prone to hypothetical bias and behavioural incongruence (Ben-Akiva and Bierlaire, 2003; Bwambale et al., 2019).

We conducted our empirical investigation in the greater Dhaka area, specifically the RAJUK area. The RP data used in our study was obtained from a secondary data source, which was originally collected for a feasibility study of the subway project in Dhaka by TYPSA (DHAKA SUBWAY - Grupo TYPSA). This dataset included information from 35,000 households and was systematically collected using stratified random sampling to represent the population characteristics of the RAJUK area. The data was collected from Monday to Saturday[16] between 28th February 2019 to 4th May 2019. The survey form had two sections: (1) general household information (e.g., age, gender, education, occupation, income, car ownership), and (2) each household member's trip-related information (e.g., departure time, travel mode, travel time, trip purpose) who made any trips during the previous working day (Sunday to Thursday). The travel diary survey recorded trips for work, education, leisure, personal and other purposes. In the case where members of a selected household declined to participate in the interview, a nearby household with a similar socio-economic profile was chosen for the interview. This often involved selecting a household located in the same building as the one that declined to be interviewed. To enhance participation, a public awareness campaign was implemented for the household interview program. This campaign included strategies such as sending text messages to approximately 13 million people to encourage their participation, along with TV scrolls. The detailed methods employed by TYPSA to ensure that this sample accurately reflected the population characteristics can be found in (TYPSA, 2019).

A total of 35,000 households were surveyed in the RAJUK Area. In this study, we concentrated only on car commuters (i.e., either privately owned, provided by the office, shared with friends/colleagues, commercial ride-hailing) and used 950 home-to-work-to-home trips. Since only a small number of individuals in the data reported multiple trips, we used one trip per person, with the earliest trips made by the commuters considered for this study. Commuting trips that had their origin outside Dhaka were not considered since that decision would be reliant on the traffic situation in the origin area. The socio-demographic characteristics of the commuters included in the sample are summarised in Table 3-1.

Table 3-1 Summary of socio-demographic characteristics of the commuters in the sample.

	Percentage
	Total respondents (n=950)
Gender	
Male	83.16
Female	16.84
Age	
<26	4.00
26 -40	37.68
40-60	47.47
>60	10.85
Monthly income	
<10,000 BDT	1.08
10,000-20,000 BDT	3.98
20,000-30,000 BDT	5.38
30,000-40,000 BDT	9.68
40,000-60,000 BDT	18.06
>60,000 BDT	61.82
Level of education	

[16] Friday is the weekly holiday in Bangladesh.

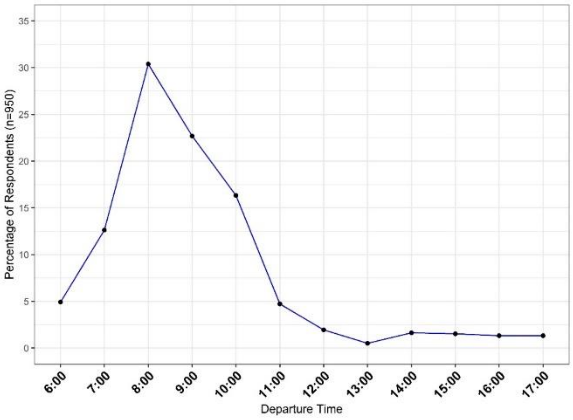
Below primary	3.81
Six to ten	5.93
SSC	5.93
HSC	9.74
BA	18.00
MA	55.03
Others	1.56
Occupation	
Public Employee	20.63
Private employee	35.58
Self-employed	36.84
Other	6.95
Car ownership rate	
No car owned by the household*	12.11
Have at least one car owned by the household	87.89

Source: DTCA, 2019

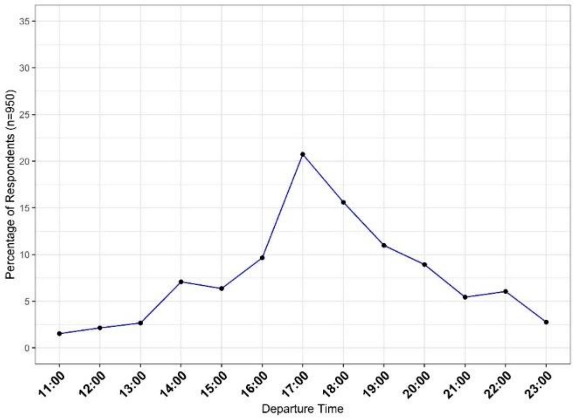
* In this case, respondents were sharing cars with friends/ colleagues or using cars provided from the office or ride-hailing services

It may be noted that although the original sample was representative of the population of Dhaka city, the sample used to develop departure time choice models in this study is likely to be biased towards high-income and educated segments of the population due to its focus on car commuters.

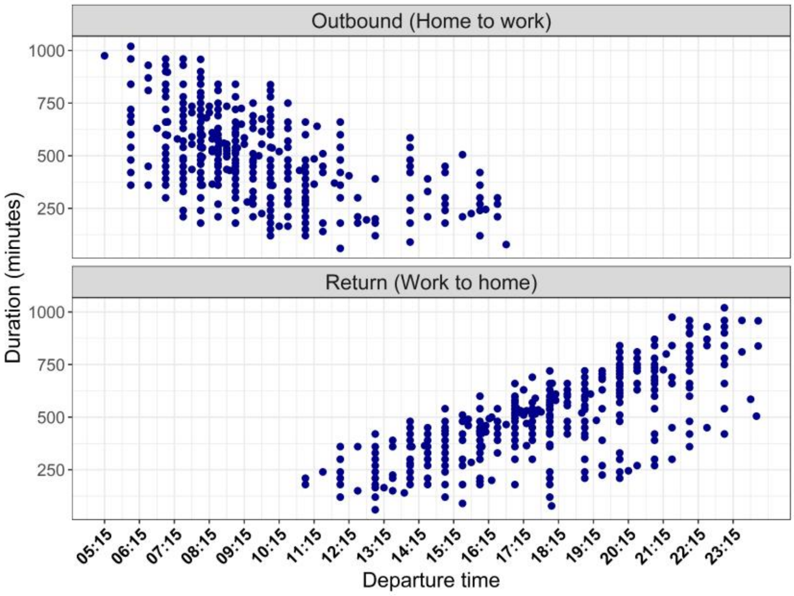
The observed departure time choices of commuters for both their outbound and return trips, along with the corresponding durations (interval between outbound departure and return), are shown in Figure 3-1 (a, b, c). The departure time distribution for the return trips exhibited a higher standard deviation compared to that of outbound trips. For the outbound trips, the peak was observed at 8:00 and for the return trips, the peak is observed at 17:00. In terms of duration, the majority of travellers experienced a difference of 6-10 hours between their outbound and return trips. The observed departure distribution indicated that for outbound trips, earlier departures were more common for longer durations, while for return trips, departures tended to be later when longer durations were required (Figure 3-1 (c)).



(a): Outbound



(b): Return



(c) Duration

Figure 3-1 Observed departure time.

Table 3-2 Summary of observed duration.

Duration window	**Percentage**
<2 hours	3.01
2 hours – 4 hours	9.76
4 hours – 6 hours	14.64
6 hours – 8 hours	26.17
8 hours – 10 hours	27.12
10 hours – 12 hours	11.32
>=12 hours	7.98

3.3. Modelling Issues

3.3.1. Choice set specification

The choice set specification is a complex step in developing a discrete choice-based departure time choice model. At this stage, the number and length of time periods are determined by subdividing the continuous time into a finite number of mutually exclusive time periods. Studies have used either a small number of the coarse time periods or a large number of fine time periods. Ben-Akiva and Bierlaire (2003) proposed a method to define the acceptable range of departure time intervals based on the preferred arrival time (PAT). However, in RP data, information related to PAT is typically not available and overestimation of the time interval may cause substantial errors. Further, in a usual specification of a joint model (simultaneous consideration of both outbound and return), a separate constant is recommended for each possible combination of home to work (outbound) and work to home (inbound) time periods. For example, *24 (N)* 1-hour separate time periods for commuting trips would lead to a requirement of 300 constants (following the rule *N(N+1)/2*), of which 299 (*N(N+1)/2-1)* can be estimated (Hess et al., 2007)). Similarly, the required number of constants would be 1,176 if 30-minute time intervals are considered for 24 hours. Thus, the increasing number of constants may lead to compounding problems of computational cost and parameter identification issues. Also, the correlation among the alternatives cannot be ignored when time intervals are short (Ben-Akiva and Bierlaire, 2003). Therefore, in this study, we selected 9 time periods for outbound (6:00 – 7:00, 7:00 –

8:00, 8:00 – 9:00, 9:00 – 10:00, 10:00 – 11:00, 11:00 – 12:00, 12:00 – 14:00, 14:00 – 16:00, 16:00-17:00) and 9 periods for return (11:00 – 12:00, 12:00 – 14:00, 14:00 – 16:00, 16:00- 17:00, 17:00 – 18:00, 18:00 – 19:00, 19:00 – 20:00, 20:00 – 22:00, 22:00 - 24:00. These time periods were divided into ten 1-hour time periods (6:00 – 7:00, 7:00 – 8:00, 8:00 – 9:00, 9:00 – 10:00, 10:00 – 11:00, 11:00 – 12:00, 16:00- 17:00, 17:00 – 18:00, 18:00 – 19:00, 19:00 – 20:00) and four 2-hour time periods (12:00 – 14:00, 14:00 – 16:00, 20:00 – 22:00, 22:00 – 24:00). We decided not to use a finer temporal resolution (e.g., 5 to 10 minutes) to avoid correlation among alternatives of a short time interval and reduce model complexity. A total of 75 alternative outbound and return combinations of choice were specified. The choice set used in this study is summarised in Table 3-3. In order to forecast the probability of unchosen alternatives, all the joint combination of alternatives shown in Table 3-3 were included in the model.

Table 3-3 Joint time periods (outbound and return) used for modelling.

ID	Outbound	Return	ID	Outbound	Return	ID	Outbound	Return	ID	Outbound	Return
1	6:00 – 7:00	11:00 – 12:00	20	8:00 – 9:00	12:00 – 14:00	39	10:00 – 11:00	14:00 – 16:00	58	12:00 – 14:00	17:00 – 18:00
2	6:00 – 7:00	12:00 – 14:00	21	8:00 – 9:00	14:00 – 16:00	40	10:00 – 11:00	16:00 – 17:00	59	12:00 – 14:00	18:00 – 19:00
3	6:00 – 7:00	14:00 – 16:00	22	8:00 – 9:00	16:00 – 17:00	41	10:00 – 11:00	17:00 – 18:00	60	12:00 – 14:00	19:00 – 20:00
4	6:00 – 7:00	16:00 – 17:00	23	8:00 – 9:00	17:00 – 18:00	42	10:00 – 11:00	18:00 – 19:00	61	12:00 – 14:00	20:00 - 22:00
5	6:00 – 7:00	17:00 – 18:00	24	8:00 – 9:00	18:00 – 19:00	43	10:00 – 11:00	19:00 – 20:00	62	12:00 – 14:00	22:00 – 24:00
6	6:00 – 7:00	18:00 – 19:00	25	8:00 – 9:00	19:00 – 20:00	44	10:00 – 11:00	20:00 - 22:00	63	14:00 – 16:00	14:00 – 16:00
7	6:00 – 7:00	19:00 – 20:00	26	8:00 – 9:00	20:00 - 22:00	45	10:00 – 11:00	22:00 – 24:00	64	14:00 – 16:00	16:00 – 17:00
8	6:00 – 7:00	20:00 - 22:00	27	8:00 – 9:00	22:00 – 24:00	46	11:00 – 12:00	11:00 – 12:00	65	14:00 – 16:00	17:00 – 18:00
9	6:00 – 7:00	22:00 – 24:00	28	9:00 – 10:00	11:00 – 12:00	47	11:00 – 12:00	12:00 – 14:00	66	14:00 – 16:00	18:00 – 19:00
10	7:00 – 8:00	11:00 – 12:00	29	9:00 – 10:00	12:00 – 14:00	48	11:00 – 12:00	14:00 – 16:00	67	14:00 – 16:00	19:00 – 20:00
11	7:00 – 8:00	12:00 – 14:00	30	9:00 – 10:00	14:00 – 16:00	49	11:00 – 12:00	16:00 – 17:00	68	14:00 – 16:00	20:00 - 22:00
12	7:00 – 8:00	14:00 – 16:00	31	9:00 – 10:00	16:00 – 17:00	50	11:00 – 12:00	17:00 – 18:00	69	14:00 – 16:00	22:00 – 24:00
13	7:00 – 8:00	16:00 – 17:00	32	9:00 – 10:00	17:00 – 18:00	51	11:00 – 12:00	18:00 – 19:00	70	16:00 – 17:00	16:00 – 17:00
14	7:00 – 8:00	17:00 – 18:00	33	9:00 – 10:00	18:00 – 19:00	52	11:00 – 12:00	19:00 – 20:00	71	16:00 – 17:00	17:00 – 18:00
15	7:00 – 8:00	18:00 – 19:00	34	9:00 – 10:00	19:00 – 20:00	53	11:00 – 12:00	20:00 - 22:00	72	16:00 – 17:00	18:00 – 19:00
16	7:00 – 8:00	19:00 – 20:00	35	9:00 – 10:00	20:00 - 22:00	54	11:00 – 12:00	22:00 – 24:00	73	16:00 – 17:00	19:00 – 20:00
17	7:00 – 8:00	20:00 - 22:00	36	9:00 – 10:00	22:00 – 24:00	55	12:00 – 14:00	12:00 – 14:00	74	16:00 – 17:00	20:00 – 22:00
18	7:00 – 8:00	22:00 – 24:00	37	10:00 – 11:00	11:00 – 12:00	56	12:00 – 14:00	14:00 – 16:00	75	16:00 – 17:00	22:00 – 24:00
19	8:00 – 9:00	11:00 – 12:00	38	10:00 – 11:00	12:00 – 14:00	57	12:00 – 14:00	16:00 – 17:00			

3.2. Factors influencing departure time choice

Departure time choices depend on multiple factors that are interrelated to each other. Earlier studies examined the influence of travellers' choice decisions as a function of transportation system characteristics and level of service attributes (e.g. travel time, travel cost), individual and household sociodemographic characteristics, activity-related attributes (e.g. mandatory vs. discretionary) (Bhat and Steed, 2002; Sasic and Habib, 2013). Findings of the previous studies show that the departure time choice of individuals is substantially affected by travel time, travel cost and travel distance which are often marked as the level of service attributes or network variables (Ben-Akiva et al., 1985; Abou-Zeid et al., 2006; Arellana et al., 2012; Zhu et al., 2018). Other studies have investigated the influence of schedule delay on departure time choice (Hess et al., 2005; Koppelman et al., 2008; Yang, G. and Liu, 2018; Bwambale et al., 2019). Other contributing factors include individual (e.g. age, gender, level of education, having a driving license, working status, flexibility at work etc.) and household attributes (e.g. household size, income, vehicle ownership, house location) (Bhat and Steed, 2002; Yang, F. et al., 2008; Arellana et al., 2013; Anowar et al., 2019; Afandizadeh Zargari and Safari, 2020; Rahman et al., 2021).

Based on the review, the level of service attributes (travel time), individual (age, gender, occupation, education) and household (household income, house location, household size, having dependant within the household) sociodemographic characteristics and trip-related attributes (available mode, distance) were considered in both the proposed and trigonometric model by Ben-Akiva and Abou-Zeid (2013).

3.3.2. Estimation of travel time

One of the key challenges to model departure time choice is the estimation of travel time during the unchosen time periods. In many cities, Google Maps and, Open Street Maps provide reliable travel times for each alternative time period with adequate spatial and temporal granularity which can be used for deriving travel times during different time periods for different origin-destination pairs (e.g. Bwambale et al. (2019) and Dong and Cirillo (2020)). But in the context of Dhaka, the widely used network traffic model of Google Maps API (i.e., best guess) does not consistently reflect a reasonable travel time that matches the users' experienced travel time. Instead, it offers three network travel times for each origin-destination pair within a given time period (best guess[17], pessimistic[18], optimistic[19]). At some time period, the travel time from the pessimistic model appears to be more closely aligned with the user-stated travel time, while in other instances, the other two models demonstrate better alignment. Therefore, we estimated the travel time for both chosen and unchosen time period following the method proposed by Zannat et al. (2022) who hypothesised that the commuter stated travel time is linearly correlated with the predicted travel time of Google Maps (direction) API. The relationship between stated travel time and the best guess, pessimistic and optimistic travel times can be expressed as follows:

$$T_{stated\ travel\ time\_i} = W_{i,1}T_{Best\ guess\_i} + W_{i,2}T_{Optimistic\_i} + W_{i,3}T_{Pessimistic\_i} + \varepsilon \qquad (1)$$

[17] Best guess model returns the duration in traffic using both historical traffic conditions and live traffic. Live traffic becomes more important the closer the departure time is to now.

[18] Pessimistic model returns the duration in traffic, usually that should be longer than the actual travel time on most days, though occasional days with particularly bad traffic conditions may exceed this value.

[19] Optimistic model returns the duration in traffic, usually that should be shorter than the actual travel time on most days, however, occasional days often with a good traffic condition could be faster than this value.

where,

i is the alternative time period ($i\epsilon n$, where n is 7 for home to work and 5 for work to home trip)

$T_{stated\ travel\ time\_i}$ = Stated travel time by the respondents at the time period i

$T_{Best\ guess\_i}$ = Measured travel time using best guess model of Google Maps API at the time period i

$T_{Optimistic\_i}$ = Measured travel time using optimistic model of Google Maps API at the time period i

$T_{Pessimistic\_i}$ = Measured travel time using pessimistic model of Google Maps API at the time period i

$W_{i,t}$ indicates the weights of measured travel time by different Google Maps models (t) for time period i, $W_{i,1}$, $W_{i,2}$, and $W_{i,3}$ were estimated assuming $\sum W_{i,t} = 1$. ε represents the error which was assumed to be normally distributed (0, σ). The relationship among $W_{i,t}$ and estimated parameters for different models (t) at time period i can be expressed as:

$$W_{i,t} = \frac{e^{\beta_{i,t}}}{e^{\beta_{i,1}} + e^{\beta_{i,2}} + e^{\beta_{i,3}}} \tag{2}$$

Following the normal distribution, equation 1 and equation 2 was used to estimate $\beta_{i,t}$ for three Google Maps models. Consequently, we calculated the weighted network travel time using the estimated weight $\hat{W}_{i,t}$ corresponding to various network models rather than relying on the network travel time from the randomly selected network model (t) using the Google Maps API. The network travel time for all alternatives was estimated using the Equation 3. Models and their corresponding weights used for the different alternative time periods are shown in Table 3-4.

$$T_{network\ travel\ time\_i} = \hat{W}_{i,1} T_{Best\ guess\_i} + \hat{W}_{i,2} T_{Optimistic\_i} + \hat{W}_{i,3} T_{Pessimistic\_i} \tag{3}$$

Table 3-4 Calculated weights of the different models used for travel time calculation.

(a) Home to Work Trip

Home to Work Trip	**Google Maps Model**	**Estimates ($\boldsymbol{\beta_{i,t}}$)**	**Exp ($\boldsymbol{\beta_{i,t}}$)**	**Weight ($\hat{W}_{i,t}$)**
		$\sigma = 31.4626$		
6:00 – 7:00	Best Guess	14.2567	1554555.173	0.9996[(a)]
	Optimistic	6.4678	644.0652237	0.0004
	Pessimistic	0	1	0.0000
		$\sum$	1555200.238	1
7:00 – 8:00	Best Guess	-15.9892	1.13757E-07	0.0000
	Optimistic	1.1985	3.315140481	0.7683
	Pessimistic	0	1	0.2317
		$\sum$	4.315140594	1
8:00 – 9:00	Best Guess	-12.567	3.48515E-06	0.0000
	Optimistic	0.239	1.269978537	0.5595
	Pessimistic	0	1	0.4405
		$\sum$	2.269982022	1
9:00 –10:00	Best Guess	-12.567	3.48515E-06	0.0000
	Optimistic	0.239	1.269978537	0.5595
	Pessimistic	0	1	0.4405

Home to Work Trip	Google Maps Model	Estimates ($\boldsymbol{\beta}_{i,t}$)	Exp ($\boldsymbol{\beta}_{i,t}$)	Weight ($\hat{W}_{i,t}$)
		$\sum$	2.269982022	1
10:00 – 11:00	Best Guess	-1.8409	0.158674555	0.0559
	Optimistic	0.5177	1.678163432	0.5916
	Pessimistic	0	1	0.3525
		$\sum$	2.836837986	1
11:00 – 12:00	Best Guess	0.0022	1.002202422	0.4997
	Optimistic	-5.6696	0.003449245	0.0017
	Pessimistic	0	1	0.4986
		$\sum$	2.005651666	1
12:00 and after 12:00	Best Guess	-0.7994	0.449598642	0.0000
	Optimistic	17.0538	25490082.17	1.0000(b)
	Pessimistic	0	1	0.0000
		$\sum$	25490083.62	1

(b) Work to Home Trip

Work to Home trip	Google Maps Model	Estimates ($\beta_{i,t}$)	Exp ($\beta_{i,t}$)	Weight ($\hat{W}_{i,t}$)
11:00 – 12:00	Best Guess	0.0022	1.002202422	0.4997
	Optimistic	-5.6696	0.003449245	0.0017
	Pessimistic	0	1	0.4986
		$\sum$	2.005651666	1
12:00 – 14:00	Best Guess	1.763	5.829900889	0.0009
	Optimistic	8.7768	6482.101226	0.9989(c)
	Pessimistic	0	1	0.0002
		$\sum$	6488.931127	1
16:00 – 18:00	Best Guess	-9.1197	0.000109488	0.0001
	Optimistic	-1.7165	0.179693977	0.1523
	Pessimistic	0	1	0.8476
		$\sum$	1.179803465	1
18:00 – 19:00	Best Guess	-5.1375	0.005872352	0.0007
	Optimistic	2.0616	7.85853341	0.8865
	Pessimistic	0	1	0.1128
		$\sum$	8.864405762	1
19:00 – 24:00	Best Guess	7.1876	1322.924374	0.9880
	Optimistic	2.7108	15.04130375	0.0112
	Pessimistic	0	1	0.0007
		$\sum$	1338.965678	1

Note: Given the close to 1 weight, only the Best Guess model has been used in (a) and only the Optimistic model has been used in (b) and (c) instead of the weighted models.

3.4. Model structure

3.4.1. Proposed model framework

The proposed model structure was based on random utility framework. Random utility theory suggests that individual decision is driven by rationality and complete information. Decision-makers choose the departure time that provides them with the highest utility, where the utility of an alternative i to a person n has the form:

$$U_{in} = U(x_{in}, s_n) \tag{4}$$

where x_{in} is the vector of the attribute of alternative i for individual n and s_n is the vector of characteristics of individual n.

McFadden (1973) proposed that this utility has the linear-in-parameters separable form:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{5}$$

where V_{in} is the observed component of utility. The unobserved variable ε_{in} expresses the contribution of unobserved attributes to the utility. In our model, ε_{in} was assumed to be independent and identically distributed across alternatives and respondents, following a Type I Extreme Value distribution (Gumbel). Therefore, the time preference of commuters was estimated using the multinomial logit model (MNL).

Further, following our proposed polynomial formulation, for an alternative time period (departure time period for outbound *o*, departure time period for return *r*), the systematic utility for the home-based commuting trip was specified as sum of three components corresponding to the outbound departure time ($V_{in}{}^{dept}$), the duration ($V_{in}{}^{dur}$) and the interaction between two ($V_{in}{}^{int}$).

$$V_{in} = V_{in}{}^{dept} + V_{in}{}^{dur} + V_{in}{}^{int} \tag{6}$$

where, $V_{in}{}^{dept}$, $V_{in}{}^{dur}$, $V_{in}{}^{int}$ can be specified as follows:

$$V_{in}{}^{dept} = \sum_{k=1}^{4} s_k\, f^{dept}(t_o) + \beta\,(TT_o) \tag{7}$$

$$V_{in}{}^{dur} = \sum_{k=1}^{4} s_k\, f^{dur}(t_r - t_o) + \beta\,(TT_r) \tag{8}$$

$$V_{in}{}^{int} = \sum_{k=1}^{4} s_k\, f^{int}(t_o)(t_r - t_o) \tag{9}$$

where, t_o and t_r are the departure times from home for outbound (midpoint of period *o*) and return respectively (midpoint of period *r*). Also, s_1=1, s_2=office employee dummy, s_3 = short distance dummy (<8 km), s_4 = High income dummy (>60,000 BDT) and TT_o and TT_r are the corresponding travel time of outbound and return. It is important to note that duration was the interval between outbound departure (t_o) and return (t_r) which included travel from home to work and activity duration.

The proposed polynomial formulations for departure time, duration and interaction can be expressed as follows:

$$f^{dept}(t_o) = \alpha_1{}^{dept} t_o + \alpha_2{}^{dept} t_o{}^2 + \cdots + \alpha_a{}^{dept} t_o{}^a \tag{10}$$

$$f^{dur}(t_r - t_o) = \alpha_1{}^{dur}(t_r - t_o) + \alpha_2{}^{dur}(t_r - t_o)^2 + \cdots + \alpha_b{}^{dur}(t_r - t_o)^b \tag{11}$$

$$f^{int}(t_o)(t_r - t_o) = \alpha_i{}^{int}(t_o)(t_r - t_o) \tag{12}$$

where a, and b are non-negative integer values defining truncation points and were determined empirically.

The unknown parameters estimated in the polynomial formulation were: $\alpha_1{}^{dept}, \ldots, \alpha_a{}^{dept}$, $\alpha_1{}^{dur}, \ldots$, $\alpha_b{}^{dur}$ and $\alpha_i{}^{int}$, for every variable s_k that was interacted with $f^{dept}(t_o)$, $f^{dur}(t_r - t_o)$ and $f^{int}(t_o)(t_r - t_o)$ in the departure time, duration and interaction component of utility function, and the travel time parameter β. However, for each variable s_k, we estimated four different interaction parameters for — 1) alternatives with peak time at both legs (outbound and return), 2) alternatives with peak time at outbound leg, 3) alternatives with peak time at return leg and, 4) alternatives with off-peak time at both legs. Such specification enabled to capture the correlation between departure time and duration, and schedule delay effect simultaneously.

The choice probabilities for each alternative i in MNL can be expressed as follows (for detail see Train (2009)):

$$P_{in} = \frac{exp\left(V_{in}{}^{dept} + V_{in}{}^{dur} + V_{in}{}^{int}\right)}{\sum_{j \in C_n} exp\left(V_{jn}{}^{dept} + V_{jn}{}^{dur} + V_{jn}{}^{int}\right)} \tag{13}$$

where, C_n is the choice set of n number of individuals (see section 3.3.1 for details).

The estimation was done using the "Apollo" package R, applying the Maximum Likelihood Estimation technique with the BFGS optimisation algorithm (Hess and Palma, 2019).

3.4.2. Trigonometric model

The trigonometric formulation proposed by Ben-Akiva and Abou-Zeid (2013) was used as the state-of-the-art model. Following this model, the systematic utility for the home-based commuting trip was specified as sum of departure time component for outbound, departure time component for return and associated duration component. Hence,

$$V_{in} = V_{in}{}^{out} + V_{in}{}^{ret} + V_{in}{}^{dur} \tag{14}$$

here, $V_{in}{}^{out}$, $V_{in}{}^{ret}$ and $V_{in}{}^{dur}$ are the outbound, return and duration component of utility.

$V_{in}{}^{out}$ and $V_{in}{}^{ret}$ were then specified as follows:

$$V_{in}{}^{out} = \sum_{k=1}^{4} s_k\, f^{out}(t_o) + \ln\,(number\ of\ hours\ in\ period\ o) + \beta\ (TT_o) \tag{15}$$

$$V_{in}{}^{ret} = \sum_{k=1}^{4} s_k\, f^{ret}(t_r) + +\ln\,(number\ of\ hours\ in\ period\ r) + \beta\ (TT_r) \tag{16}$$

The trigonometric formulation for outbound and return can be expressed as follows:

$$\begin{aligned} f^{out}(t_o) = {} & \alpha_1{}^{out} \sin\left(\frac{2\pi t_o}{24}\right) + \alpha_2{}^{out} \sin\left(\frac{4\pi t_o}{24}\right) + \alpha_3{}^{out} \sin\left(\frac{6\pi t_o}{24}\right) + \alpha_4{}^{out} \sin\left(\frac{8\pi t_o}{24}\right) \\ & + \alpha_5{}^{out} \cos\left(\frac{2\pi t_o}{24}\right) + \alpha_6{}^{out} \cos\left(\frac{4\pi t_o}{24}\right) + \alpha_7{}^{out} \cos\left(\frac{6\pi t_o}{24}\right) \\ & + \alpha_8{}^{out} \cos\,(\frac{8\pi t_o}{24}) \end{aligned} \tag{17}$$

$$f^{ret}(t_r) = \alpha_1{}^{ret} \sin\left(\frac{2\pi t_r}{24}\right) + \alpha_2{}^{ret} \sin\left(\frac{4\pi t_r}{24}\right) + \alpha_3{}^{ret} \sin\left(\frac{6\pi t_r}{24}\right) + \alpha_4{}^{ret} \sin\left(\frac{8\pi t_r}{24}\right) \\ + \alpha_5{}^{ret} \cos\left(\frac{2\pi t_r}{24}\right) + \alpha_6{}^{ret} \cos\left(\frac{4\pi t_r}{24}\right) + \alpha_7{}^{ret} \cos\left(\frac{6\pi t_r}{24}\right) + \alpha_8{}^{ret} \cos\left(\frac{8\pi t_r}{24}\right) \quad (18)$$

The duration component $V_{in}{}^{dur}$ can be specified as a power series expansion as follows:

$$V_{in}{}^{dur} = \alpha_1{}^{dur}(t_r - t_o) + \alpha_2{}^{dur}(t_r - t_o)^2 + \cdots + \alpha_b{}^{dur}(t_r - t_o)^b \quad (19)$$

The unknown parameters estimated in the trigonometric formulation were: $\alpha_1{}^{out}, \ldots, \alpha_8{}^{out}$ and $\alpha_1{}^{ret}, \ldots, \alpha_8{}^{ret}$ for every variable s_k interacted with $f^{out}(t_o)$, $f^{ret}(t_r)$ in the utility function, and the travel time parameter β. From the power expansion of duration utility, the estimated parameters were $\alpha_1{}^{dur}, \ldots, \alpha_b{}^{dur}$. Here, the b value depends on empirical consideration. Also, the size variable was included for the outbound and return departure time period. Similar to our proposed model, the coefficients were estimated using the Maximum Likelihood Estimation technique.

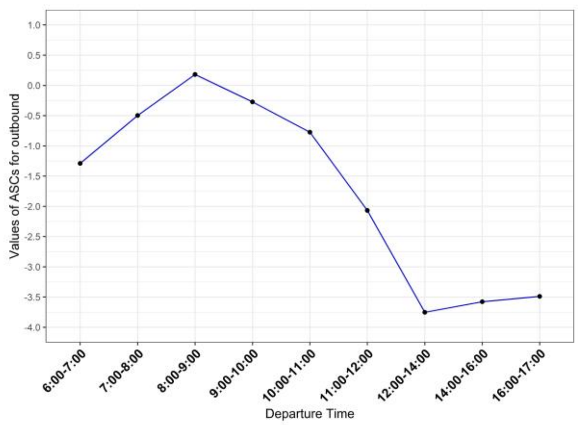
3.5. Results and discussion

3.5.1. Base MNL Models

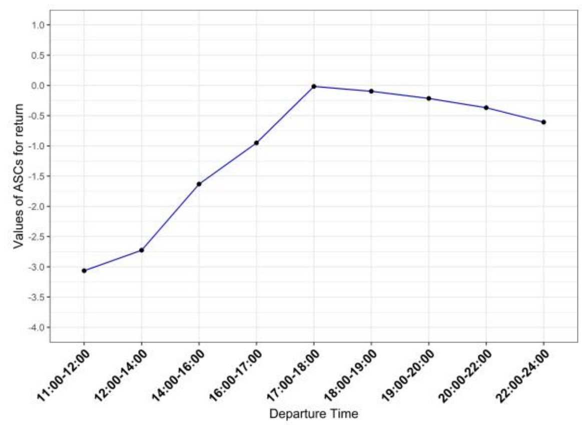
Table 3-5 shows the estimates of the simple MNL models fitted to the RP data developed using the state-of-the-art method and proposed polynomial formulation. These two base models were used to investigate the performance and interpretation capabilities of the proposed polynomial functional form. The models included travel time as the main explanatory variable, while other covariates (e.g., socio-demographic characteristics) were excluded. When comparing the adjusted Rho-square, and BIC, the results showed that the polynomial model produced a slightly better model fit than the trigonometric model. Additionally, the proposed model's RMSE value demonstrated that it had less error than the trigonometric model. This study also revealed that the polynomial approximation performed better than the trigonometric model when evaluating how travel time affected decisions. In the trigonometric model as opposed to the proposed model, the estimated marginal utility of travel time was lower. Also, in the proposed model, the interaction parameters were found to be statistically significant in modelling the time choice of travellers (the null hypothesis of zero correlations between departure times and durations was rejected at a 95% level of confidence). The estimated coefficients of interaction parameters demonstrated that, within a time budget of 24 hours per day, if car travellers chose a longer duration (including travel time to workplace and activity duration), they were less likely to prefer a later time of the day for departing from home. Moreover, the relatively lower interaction effect on those alternatives having peak time in both legs (outbound and return) reflected the lower schedule delay compared to the other alternatives.

The inferred shape of the ASCs' value from the trigonometric formulation (Figure 3-2 (a) and Figure 3-2 (b)) indicated potential overfitting of the models during the afternoon for the outbound and late evening for return journeys. This is due to the fact that late night alternatives for returning were not as popular as the afternoon and morning alternatives, respectively, based on the observation. Similarly, for outbound, the evening was the least popular than the other alternatives (Figure 3-1). On the other hand, the aggregated utility of the duration component (estimated using power expansion) of the state-of-the-art method revealed that the highest utility was at the duration window between 2-4 hours (Figure 3-2 (c)). But the observation showed that the highest percentage of respondents chose the duration between 8-10 hours (Table 3-2).

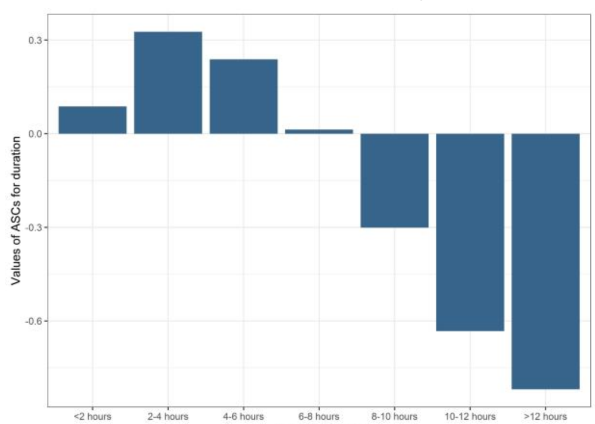
Figure 3-3 shows a surface plot derived from the proposed polynomial formulation, to investigate the influence of outbound, duration, and corresponding correlation, all together. All else being equal, the combination of time choices for outbound and return journeys were ranked as follows according to the level of preference: (1) 9:00 – 10:00 and 18:00 – 19:00, (2) 9:00 – 10:00 and 17:00 – 18:00, and (3) 8:00 – 9:00 and 17:00 – 18:00. Such results indicated that the utility of departure from home was higher between 8:00 – 10:00 and 17:00 – 19:00 for departure from work, which complied with the observed departure times (Figure 3-1 (a) and Figure 3-1 (b)). At the same time, the intrinsic preference of duration was greater than 9 hours (estimated from the mid-point difference of return and outbound alternatives) which included the travel time to the workplace and activity duration. This result is intuitive because, in Bangladesh, the earliest office starting time is between 9:00 to 10:00, and closing time is between 16:00 to 17:00 which gives an average of 8 working hours (including lunch break). Therefore, from the base model, we can infer that our results complied with reality and the proposed model was appropriately fitted with the RP data of car commuters both in the morning and evening.



(a) Intrinsic preference of departure time of outbound car commuters while controlling travel time sensitivity.



(b) Intrinsic preference of departure time of return car commuters while controlling travel time sensitivity.



(c) Intrinsic preference of duration while controlling travel time sensitivity.

Figure 3-2 Values of ASCs for (a) outbound, (b) return, and (c) duration following trigonometric formulation.

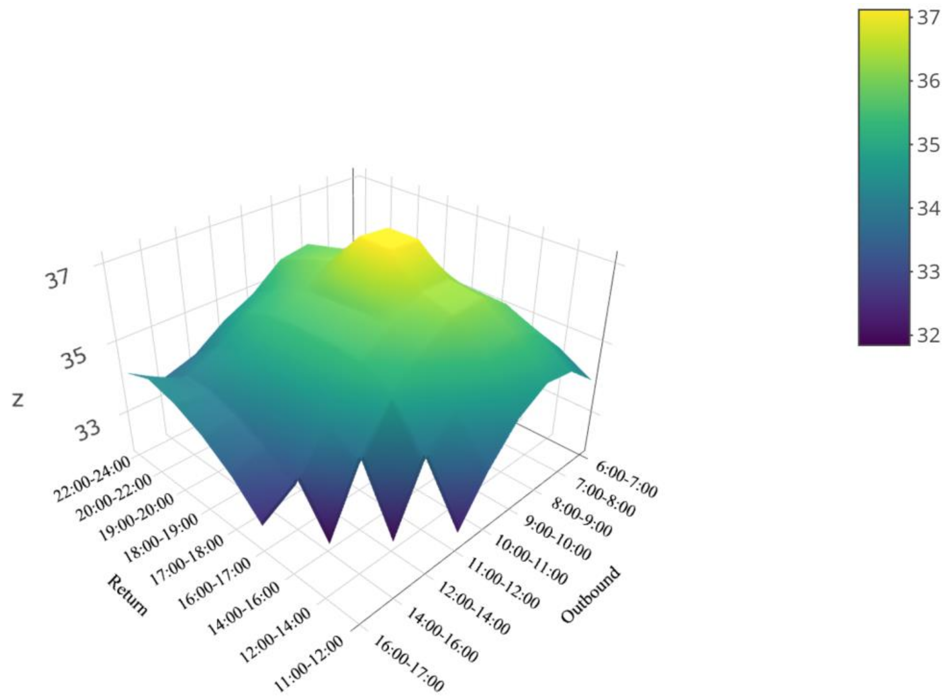


Figure 3-3 Values of ASCs following the polynomial formulation while controlling travel time sensitivity (Highest utility at outbound 8:00 – 10:00 and return 17:00 – 19:00 combination).

Table 3-5 Estimates from the base MNL model.

(a) State-of-the-art model			**(b) Proposed model**		
Parameter	**Estimate**	**Rob.t.rat.(0)**	**Parameter**	**Estimate**	**Rob.t.rat.(0)**
${\alpha_1}^{out}$	3.025	8.732	${\alpha_1}^{dept}$	8.210	8.124
${\alpha_2}^{out}$	3.257	7.223	${\alpha_2}^{dept}$	-0.683	-7.647
${\alpha_3}^{out}$	-1.740	-4.694	${\alpha_3}^{dept}$	0.018	7.174
${\alpha_4}^{out}$	-1.729	-6.543	${\alpha_1}^{dur}$	1.748	8.555
${\alpha_5}^{out}$	0.970	5.045	${\alpha_2}^{dur}$	-0.126	-6.869
${\alpha_6}^{out}$	-1.970	-10.403	${\alpha_3}^{dur}$	0.004	5.269
${\alpha_7}^{out}$	-2.440	-6.874	$\alpha^{int} out\,\&\,ret\ offpeak$	-0.046	-4.761
${\alpha_8}^{out}$	-1.016	-5.470	$\alpha^{int} out\ peak$	-0.043	-4.392
${\alpha_1}^{ret}$	-2.488	-3.424	$\alpha^{int} ret\ peak$	-0.047	-4.582
${\alpha_2}^{ret}$	2.291	4.206	$\alpha^{int} out\,\&\,ret\ peak$	-0.034	-3.333
${\alpha_3}^{ret}$	-0.319	-2.678	$\boldsymbol{\beta_{TT}}$	-0.012	-3.707
${\alpha_4}^{ret}$	-1.233	-4.049			
${\alpha_5}^{ret}$	-1.197	-7.084			
${\alpha_6}^{ret}$	-1.657	-7.118			
${\alpha_7}^{ret}$	1.878	4.381			
${\alpha_8}^{ret}$	-1.300	-2.658			
${\alpha_1}^{dur}$	0.642	3.064			

(a) State-of-the-art model			(b) Proposed model		
Parameter	Estimate	Rob.t.rat.(0)	Parameter	Estimate	Rob.t.rat.(0)
${\alpha_2}^{dur}$	-0.092	-4.151			
${\alpha_3}^{dur}$	0.003	3.665			
$\boldsymbol{\beta_{TT}}$	**-0.009**	**-2.515**			
LL (0)		-4101.61	LL (0)		-4101.61
LL (final)		-3651.42	LL (final)		-3655.72
Rho-square (0)		0.1098	Rho-square (0)		0.1087
Adj.Rho-square		0.1049	Adj.Rho-square		0.106
AIC		7342.84	AIC		7333.43
BIC		7439.97	BIC		7386.86
RMSE		5.35	RMSE		4.58
Estimated parameters		20	Estimated parameters		11

3.5.2. Comprehensive MNL models with sociodemographic factors

Separate MNL models were developed using both the trigonometric model by Ben-Akiva and Abou-Zeid (2013) and the proposed polynomial formulation where base models interacted with socio-demographic variables. Three different socio-demographic factors were considered using dummy variables that interacted with the constants of the base MNL models: occupation, trip-related attributes (e.g., trip length), and household income level. The resultant performance assessment indicators and estimates of both models are summarised in Table 3-6 and Table 3-7, respectively. The inclusion of socio-demographic variables with the base model led to a significant gain over the base model. The log-likelihood of trigonometric formulation increased by 121.46 over the base MNL model while incorporating 34 additional parameters for sociodemographic factors. On the other hand, the log-likelihood of the proposed polynomial formulation increased by 110.68 over the base MNL model with 24 additional parameters. The log-likelihood ratio test (LR) shows a significant improvement compared to the base MNL model (Table 3-6). However, the significance of the proposed formulation was distinguishable while comparing the adjusted rho-square, BIC, and RMSE of both models (Table 3-6). The RMSE of the proposed model was lower than the RMSE of the trigonometric model, showing that the proposed polynomial model had less error than the trigonometric model. The intrinsic preference of departure time for different socio-demographic groups is shown graphically for both the state-of-the-art model and the proposed model (Figure 3-4 and Figure 3-5).

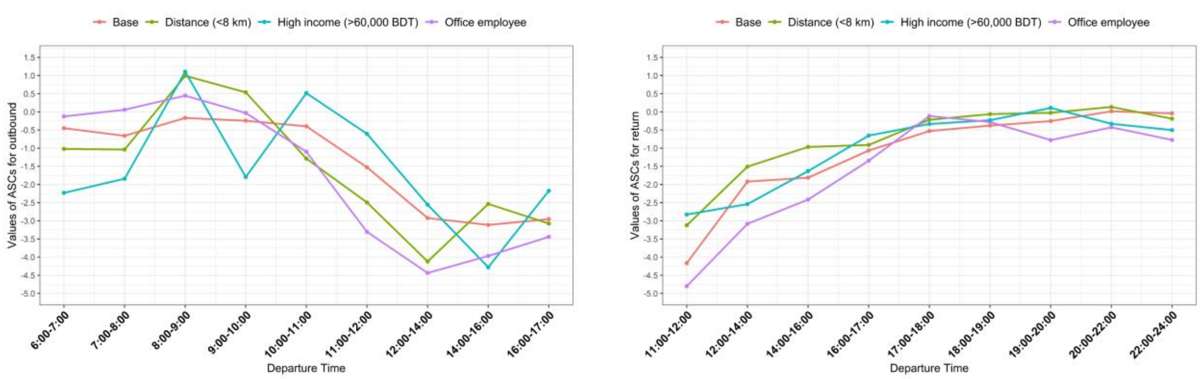
Table 3-6 Performance statistics of comprehensive MNL models.

Model Parameters	(a) State-of-the-art model	(b) Proposed model
Number of observations	950	950
Number of estimated parameters	54	35
LL (0)	-4101.61	-4101.61
LL (final)	-3529.96	-3545.04
Rho-square (0)	0.1374	0.1357
Adj.Rho-square	0.1262	0.1272
AIC	7167.92	7160.08
BIC	7430.17	7330.05

RMSE	4.70	4.43
LR test result Model with sociodemographic variables vs. base model		
Likelihood ratio test-value (c^2)	242.92	221.36
Degrees of freedom	34	24
P-value	2.196e-33	7.259e-34

3.5.2.1. Results from the model developed using trigonometric model

Results of the trigonometric model are presented in Figure 3-4 (a), Figure 3-4 (b), and Table 3-7 (a). Figure 3-4 (a) shows that the utility of outbound travel for all car commuters was larger between 7:00 and 10:00. Compared to the self-employed travellers, the office employees had higher utilities of departure for outbound trips between 8:00 – 9:00, followed by 7:00 – 8:00. The highest utility of departure for outbound trips of short distance travellers was between 8:00 – 10:00 compared to the long-distance travellers. The high-income households had the higher utility during the morning peak (8:00 – 9:00) and a later time in the morning (10:00 – 11:00) for outbound trips compared to the commuters with a monthly household income below 60,000 BDT. Figure 3-4 (b) shows the intrinsic preferences of departure times for return trips among different socio-demographic groups formulated using the estimated values of ASCs using trigonometric formulation (while controlling travel time sensitivity). The return trips were less heterogeneous across different groups. The preference of departure time was similar for short-distance car commuters and commuters from high-income households, while the pattern of the preferences of office employees was different. For office employees, the utility of departure for return trips was higher between 17:00 – 18:00 than that of self-employed commuters. However, other groups such as short-distance travellers and high income had a higher utility of departure for return after 19:00.



(a) Outbound (b) Return

Figure 3-4 Intrinsic preference of departure for different market shares while controlling travel time sensitivity.

*These plots are prepared based on values of ASCs

3.5.2.2. Results from the model developed using the proposed polynomial formulation

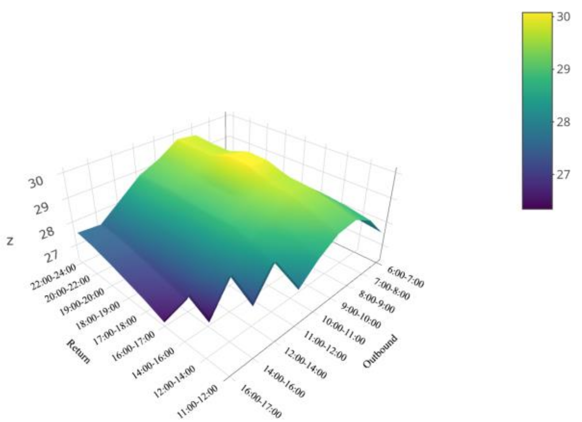
Results from the model developed using the polynomial formulation are shown in Table 3-7 (b). Surface plots in Figure 3-5 (a-d) show the preference of outbound, return, and duration for different socio-demographic groups while controlling travel time sensitivity. The results indicated that the interaction coefficients of outbound and duration varied across different socio-demographic groups. For example, office employees and people from high-income households were more sensitive to interaction effects compared to other groups. The inclusion of interaction parameters for different socio-demographics nulled the significance of the base interaction parameters, highlighting the heterogeneity in interaction effects among the different socio-demographic groups.

The base surface plot (Figure 3-5 (a)) indicates that, for the reference group (e.g., self-employed personnel, long distance commuters, and respondents from high-income households), the preferred time choice for the car commuters was 9:00 – 10:00 and 18:00 – 19:00, followed by 9:00 – 10:00 and 17:00 - 18:00, 8:00 – 9:00 and 17:00 - 18:00, 8:00 – 9:00 and 18:00 - 19:00. In terms of activity duration, the preferred duration choice varied from 8 to 10 hours (includes activity duration and travel time to work). More than 46% of car commuters (among the selected respondents) chose a duration window of greater than 8 hours (Table 3-2).

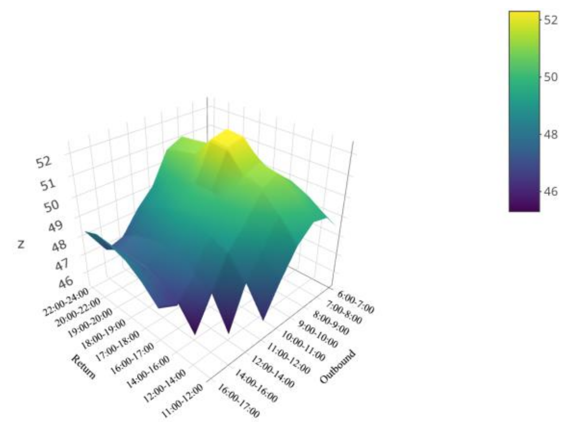
The utility of departure from home and work for the office commuters was higher before 9:00 and after 17:00, respectively, compared to other commuters (Figure 3-5 (b)). Such a scenario was also found in the trigonometric model. As mentioned, the usual office starting and closing times in Bangladesh are from 9:00 – 10:00 and 16:00 – 17:00, respectively. These periods are also known to be morning and evening peak hours. Schedule delay within these periods is expected to be minimal for the office employees, but with a greater possibility to get a late arrival penalty. Hence, to avoid the penalty of a late arrival in the morning and evening peak hours, office employees tend to prefer the other alternative periods to depart from home and work. For outbound trips of office commuters, the proposed model found the highest utility at 8:00 – 9:00; similarly, the highest utility in the trigonometric model was during 8:00 – 9:00. The observation data indicated that the highest percentage of office employees (car commuters) started their outbound journey at 8:00 – 9:00 (Supplementary Figure 3-1 (a)). Additionally, results from the polynomial model revealed that the highest duration preference of office commuters was between 9 to 10 hours, with the highest utilities for outbound and return journeys at 8:00 - 9:00 and 18:00 – 19:00, followed by 8:00 - 9:00 and 17:00 – 18:00, respectively. Since additional information about work flexibility and job type (full-time/part-time) was not available, no other additional experiment was carried out considering respondents' job types.

Furthermore, Figure 3-5 (d) exhibits that short-distance travellers (<8km) were less likely to choose a limited number of alternatives compared to long-distance travellers. Though the intrinsic preference of departure time for outbound journeys of short-distance travellers was between 8:00 – 10:00, their departure time of return journeys was distributed from 17:00 to 24:00. Such a result from the proposed model agreed with the observed data (Supplementary Figure 3-1 (c and d)). Similarly, according to the state-of-the-art model, the distribution of departure time of car commuters for outbound was relatively skewered than the distribution for return. Hence, in Dhaka where major urban roads remain congested most of the time, congestion impacts on long-distance travellers are higher than the short distance travellers.

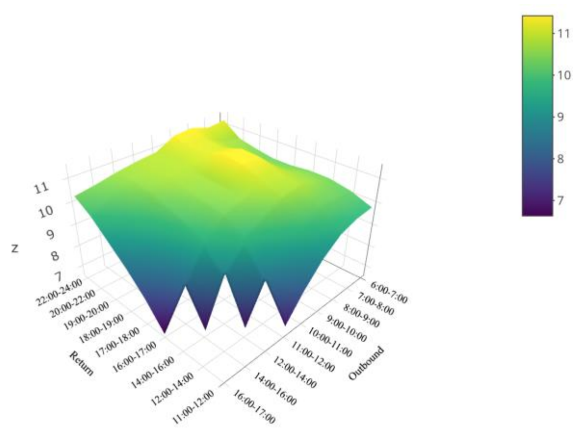
For the commuters from high-income groups (monthly income >60,000 BDT), the utility of departure time for outbound and return journeys was the highest during the morning peak (9:00 – 10:00) and evening peak (17:00 – 18:00) compared to other alternatives (Figure 3-5 (d)). A higher monthly income corresponds to a higher position in the corporate hierarchy, with less accountability for their actions. Therefore, such commuters are less likely to be affected by the consequences of a schedule delay and could prefer to travel during peak time with a very minimal effect on their schedule. For the high-income commuters, the trigonometric formulation encountered an overfitting problem as the highest utility of return was predicted at 19:00 to 20:00. However, the observation data indicated that a large share of high-income commuters travelled between 17:00 – 18:00 (Supplementary Figure 3-1 (f)).



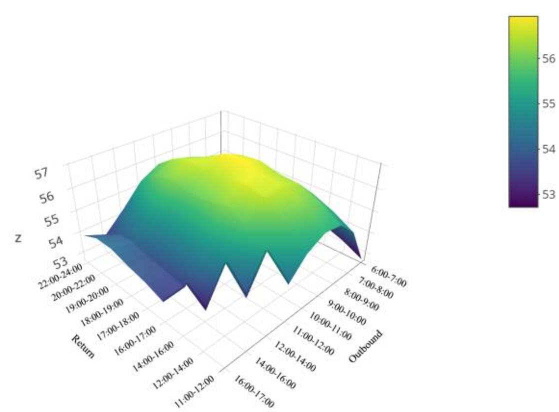
(a) Alternative specific constants for reference group (highest utility at 9:00 – 10:00 to 18:00 – 19:00)



(b) Intrinsic preference of departure for office employee compared to the self-employed personnel (highest utility at 8:00 -9:00 to 18:00 – 19:00)



(c) Intrinsic preference of departure for short distance commuters (<8km) compared to the long-distance commuters (highest utility at 8:00 -9:00 to 18:00 – 19:00)



(d) Intrinsic preference of departure for commuters from high income households compared to other income groups (highest utility at 9:00 -10:00 to 17:00 – 18:00)

Figure 3-5 Heterogeneity in preference of departure among different socio-demographic groups while controlling travel time sensitivity.

*These plots are prepared based on values of ASCs

Table 3-7 Estimates from comprehensive MNL model.

(a) State-of-the-art model			(b) Polynomial formulation		
Parameter	Estimate	Rob.t.rat.(0)	Parameter	Estimate	Rob.t.rat.(0)
$\alpha_1{}^{out}$	2.393	5.762	$\alpha_1{}^{dept}$	7.508	3.555
$\alpha_2{}^{out}$	2.889	5.887	$\alpha_2{}^{dept}$	-0.640	-3.445
$\alpha_3{}^{out}$	-1.443	-3.477	$\alpha_3{}^{dept}$	0.017	3.208
$\alpha_4{}^{out}$	-1.280	-3.740	$\alpha_1{}^{dur}$	0.416	1.422
$\alpha_5{}^{out}$	0.585	2.192	$\alpha_2{}^{dur}$	-0.039	-1.53
$\alpha_6{}^{out}$	-1.330	-5.542	$\alpha_3{}^{dur}$	0.001	1.221
$\alpha_7{}^{out}$	-2.051	-5.038	$\alpha^{int} out \,\&\, ret\ offpeak$	-0.002	-0.132
$\alpha_8{}^{out}$	-0.886	-3.032	$\alpha^{int} out\ peak$	6.7493e-04	0.041
$\alpha_1{}^{ret}$	-2.794	-2.999	$\alpha^{int} ret\ peak$	-0.003	-0.195
$\alpha_2{}^{ret}$	2.549	3.676	$\alpha^{int} out \,\&\, ret\ peak$	0.004	0.245
$\alpha_3{}^{ret}$	-0.360	-2.871	$\alpha_1{}^{dept*}$ s_2	5.372	2.539
$\alpha_4{}^{ret}$	-1.029	-2.722	$\alpha_2{}^{dept*}$ s_2	-0.507	-2.702
$\alpha_5{}^{ret}$	-0.795	-3.614	$\alpha_3{}^{dept*}$ s_2	0.015	2.875
$\alpha_6{}^{ret}$	-1.384	-4.772	$\alpha_1{}^{dur*}$ s_2	1.283	3.206
$\alpha_7{}^{ret}$	2.307	4.138	$\alpha_2{}^{dur*}$ s_2	-0.080	-2.353
$\alpha_8{}^{ret}$	-1.848	-2.922	$\alpha_3{}^{dur*}$ s_2	0.002	1.730
$\alpha_2{}^{out*}$ s_2	0.690	4.166	$\alpha^{int} out \,\&\, ret\ offpeak*$ s_2	-0.049	-2.404
$\alpha_3{}^{out*}$ s_2	-1.359	-8.122	$\alpha^{int} out\ peak*$ s_2	-0.046	-2.210
$\alpha_4{}^{out*}$ s_2	0.505	2.729	$\alpha^{int} ret\ peak*$ s_2	-0.045	-2.124
$\alpha_5{}^{out*}$ s_2	-0.610	-2.684	$\alpha^{int} out \,\&\, ret\ peak$ s_2	-0.038	-1.777
$\alpha_8{}^{out*}$ s_2	-1.209	-3.185	$\alpha_1{}^{dept*}$ s_3	-5.766	-3.845
$\alpha_1{}^{ret*}$ s_2	-0.627	-2.775	$\alpha_2{}^{dept*}$ s_3	0.499	3.625
$\alpha_2{}^{ret*}$ s_2	0.672	3.463	$\alpha_3{}^{dept*}$ s_3	-0.013	-3.322
$\alpha_4{}^{ret*}$ s_2	-0.613	-3.293	$\alpha_1{}^{dur*}$ s_3	0.768	2.686
$\alpha_5{}^{ret*}$ s_2	-0.538	-3.807	$\alpha_2{}^{dur*}$ s_3	-0.075	-2.186
$\alpha_6{}^{ret*}$ s_2	-0.900	-5.592	$\alpha_3{}^{dur*}$ s_3	0.002	1.935
$\alpha_7{}^{ret*}$ s_2	0.634	2.762	$\alpha_1{}^{dept*}$ s_4	6.074	2.884

(a) State-of-the-art model			(b) Polynomial formulation		
Parameter	**Estimate**	**Rob.t.rat.(0)**	**Parameter**	**Estimate**	**Rob.t.rat.(0)**
$\alpha_1{}^{out}*$ s_3	0.242	1.559	$\alpha_2{}^{dept}*$ s_4	-0.465	-2.497
$\alpha_4{}^{out}*$ s_3	-0.213	-1.495	$\alpha_3{}^{dept}*$ s_4	0.012	2.217
$\alpha_5{}^{out}*$ s_3	0.121	0.919	$\alpha_1{}^{dur}*$ s_4	0.838	3.130
$\alpha_6{}^{out}*$ s_3	-0.342	-3.220	$\alpha_2{}^{dur}*$ s_4	-0.026	-3.780
$\alpha_1{}^{ret}*$ s_3	1.024	3.114	$\alpha^{int} out\ \&\ ret\ offpeak*$ s_4	-0.047	-2.537
$\alpha_2{}^{ret}*$ s_3	-0.810	-3.006	$(\alpha^{int} out\ peak \mid (\alpha^{int} out\ \&\ ret\ peak))*$ s_4	-0.054	-2.767
$\alpha_4{}^{ret}*$ s_3	0.505	2.722	$\alpha^{int} ret\ peak*$ s_4	-0.050	-2.517
$\alpha_5{}^{ret}*$ s_3	0.246	1.640	$\boldsymbol{\beta_{TT}}$	**-0.013**	-4.037
$\alpha_6{}^{ret}*$ s_3	0.238	1.245			
$\alpha_7{}^{ret}*$ s_3	-0.872	-2.873			
$\alpha_8{}^{ret}*$ s_3	0.589	2.409			
$\alpha_1{}^{out}*$ s_4	0.886	3.877			
$\alpha_3{}^{out}*$ s_4	0.819	4.828			
$\alpha_4{}^{out}*$ s_4	-0.980	-3.319			
$\alpha_5{}^{out}*$ s_4	1.115	3.589			
$\alpha_6{}^{out}*$ s_4	-0.970	-3.798			
$\alpha_7{}^{out}*$ s_4	-0.590	-2.620			
$\alpha_8{}^{out}*$ s_4	0.553	1.483			
$\alpha_1{}^{ret}*$ s_4	0.524	2.200			
$\alpha_2{}^{ret}*$ s_4	-0.444	-2.348			
$\alpha_5{}^{ret}*$ s_4	-0.423	-3.490			
$\alpha_7{}^{ret}*$ s_4	-0.544	-2.779			
$\alpha_8{}^{ret}*$ s_4	0.466	2.254			
$\alpha_1{}^{dur}$	0.683	3.312			
$\alpha_2{}^{dur}$	-0.096	-4.580			
$\alpha_3{}^{dur}$	0.003	4.125			
$\boldsymbol{\beta_{TT}}$	-0.012	-3.336			

3.5.3. Benefits of the proposed polynomial formulation over the state-of-the-art method

The previous departure time choice models attempted to address different methodological issues and provided functional approaches to handle the associated modelling complexities. This study proposed a novel functional approximation that captures the correlation between departure time and duration. The results suggested that interactions between departure time and duration significantly contribute to estimating the utility of time preference. The interaction parameters were significant in formulating the utility of different market shares such as office employees and high income (Table 3-7 (b)). On the other hand, the trigonometric model did not consider the correlation between departure time and duration, which caused an overestimation of utility for these market segments. Moreover, while the trigonometric formulation captures cyclical effects, the polynomial formulation can capture a wide range of non-linear relationship between variables and offers flexibility by accommodating various distribution patterns, such as linear, cubic, or quadratic.

Besides, the proposed polynomial functional form was more flexible and required fewer parameters compared to the state-of-the-art method. As a result, the proposed polynomial functional form is computationally less expensive and does not have complex identification issues. The results of the study reinforced the finding from the previous studies, which demonstrated the level of service attributes, trip attributes, and socio-demographic factors significantly influence time-of-day choice (Hess et al., 2005; Ben-Akiva and Abou-Zeid, 2013; Bwambale et al., 2019; Palma et al., 2021; Zannat et al., 2022). This study noted that missing information (e.g., preferred activity duration, preferred arrival or departure time) in RP data caused difficulties in estimating the time-of-day choice. In such a case, overlooking the correlation between departure time and duration could affect the role of critical explanatory variables (e.g., travel time). For example, in the proposed model the use of interaction terms in the estimation process enabled to capture the larger effect of different independent factors (travel time) and dimensions (departure time and duration) of the systematic utility while comparing to the sum of the individual dimension. Also, the model used functional approximation instead of a full set of constants for alternatives using RP data. Eventually, the novelty of the proposed functional form stands on its powerful capacity to capture the heterogeneity associated with the utility of departure from home at the same time but for different durations.

3.6. Policy insights

To address congestion issues such as congestion in peak time, various studies have emphasised the necessity for incentives (reduced fare for public transport, congestion tax for private vehicles) that encourage changes in transportation modes, destinations, and departure times (Marshall and Banister, 2000; Kockelman and Kalmanje, 2005; Moya-Gómez and García-Palomares, 2017). In Dhaka, where congestion reduction strategies are still in their infancy, primarily focusing on expanding capacity and encountering challenges during peak commuting hours, options such as relocating office locations or adopting remote work practices remain unpopular (Jamal et al., 2022). Additionally, other studies have demonstrated that car users exhibit strong resistance to mode switching due to factors like comfort and time sensitivity (Khan et al., 2011). Given these constraints, adjusting departure times emerges as a viable option for mitigating peak-hour traffic congestion in major congested areas.

To gain insights into departure time preferences, it is crucial to employ a modelling framework capable of capturing the sensitivity to various aspects of time preference. In this article, we introduced a novel modelling framework that addresses a level of complexity that was previously unexplored in departure time choice modelling. The implications of our proposed framework and results can be understood in two distinct ways. Firstly, our findings highlighted a significant correlation between departure time and duration, shedding light on its importance in understanding time-of-day choices. This methodological contribution has the potential to enhance our comprehension of travellers' decision-making processes which can lead to a better policy intervention. Furthermore, the estimated model parameters have practical applications in formulating policies aimed at spreading peak-hour traffic to reduce congestion caused by car commuters in Dhaka, Bangladesh.

To highlight the practical application of the proposed model, we carried out 3 different forecasting exercises, each involving a modification of a specific attribute influencing the time-of-day choice, as considered within the model. In the first scenario, we assumed that everybody in the sample would behave as an office employee using a car for commuting. In the second scenario, it was assumed everyone in the sample would be short-distance car traveller and in the third scenario everyone would be from a high-income household. The summary of the forecasting is summarised in Figure 6. For each scenario, we presented the forecasted average probability for each time choice, along with the percentage change from the base scenario. The base prediction presented the choice context replicated by the proposed model and this prediction served as initial reference point for the forecasting. Furthermore, we presented in Table S1 the proportion of the sample that selected each alternative, in conjunction with the base prediction providing a validation of the model. In the first scenario (all car commuters office employee), a discernible rise in the likelihood of choosing morning and afternoon peaks, as opposed to other off-peak time periods, was observed. This shift was accompanied by a reduction of time choice during off-peak hours. A similar increase in peak-hour travel demand was observed when examining an increase in respondents from high-income households. However, their likelihood of shifting towards the peak hour was relatively lower compared to office employees. For these socio-demographic groups, the correlation between departure time and duration exerted a significant and dominant influence on the choice of earlier times of the day (around the time of morning and afternoon peak), especially when there was a longer duration requirement. In scenario 2, where a correlation between departure and duration was lacking, substantial shifts were noted during the morning and evening off-peaks.

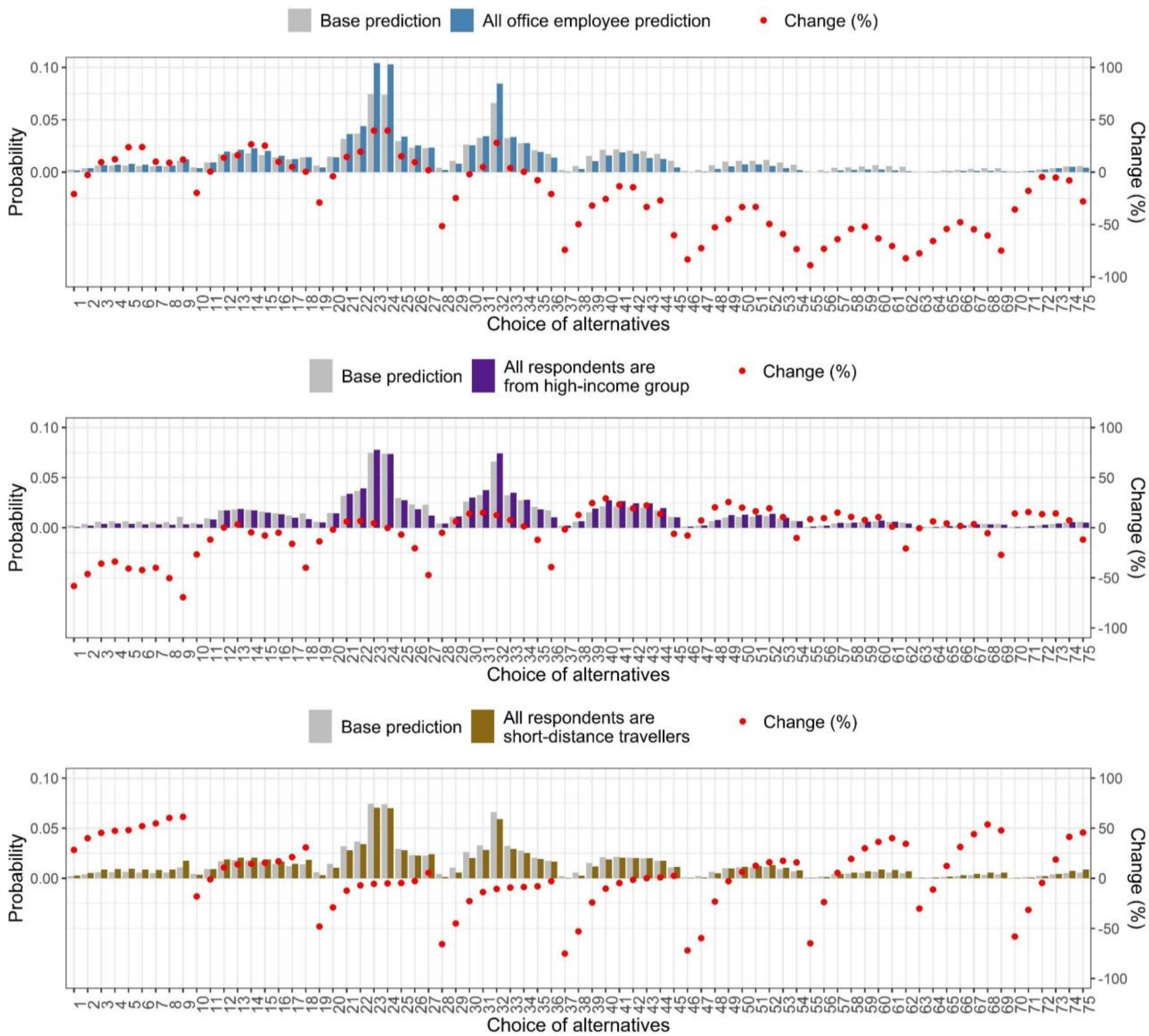


Figure 3-6 Forecasting scenario results (details of choice of alternatives can be found in Table 3.3).

Based on our results and forecasting analysis, we outline different proposals for our case study (Table 3-8).

Table 3-8 Significance of direct model output to formulate peak-spreading policies targeting car commuters.

Estimated parameters	Outcomes	Policy implication
Correlation between departure time and duration (activity and travel combined)	We found a negative correlation between departure time and duration, indicating that long duration requirements led to a preference for earlier departure times.	To meet the needs of car commuters with requirements of longer durations, it is essential to recognise that they often need to begin their journeys early in the morning. particularly in morning peaks, to ensure they have enough time to meet their workplace requirements within standard opening and closing hours. We observed that a significant portion of car commuters, approximately 46% of the respondents we selected, fall into this category,

Estimated parameters	Outcomes	Policy implication
		requiring a duration window of more than 8 hours. To alleviate the demand during the morning rush hour for those who need to travel early for longer durations, it is vital to introduce flexibility in work hours and office starting times, whenever possible. This approach strengthens the findings of a study by Kockelman and Kalmanje (2005), which highlighted that congestion pricing policies face challenges when fixed office hour requirements are in place.
Time dependent correlation	During the morning and evening peak hours, the correlation between departure time and duration appeared to be relatively weak compared to other time periods.	Results highlighted that there was a weaker association between departure time and duration during peak hours. This implies that at the morning and evening peak, the need for a longer duration had less impact on encouraging people to opt for an earlier departure before the peak hours. The potential reason can be the disutility of arriving early or late at work (Hendrickson and Plank, 1984; De Palma et al., 1990). To complement this, peak spreading policies could promote off-peak travel by offering pricing incentives for travelling before or after peak hours. Additionally, implementing flexible work schedules, such as starting work before or after the usual office hours, can further mitigate the impact of early arrival waiting time.
Effects of occupation type (office employees vs. self-employed personnel)	Outbound and return office employees preferred morning and evening peak with a longer duration requirement. Also, the forecasting analysis shows that if there were an increase in car commuter office employees, there would be a shift from other off-peak alternatives to peak periods.	The significant difference between the office employees and self-employed individuals highlighted the fixed work hours and strong schedule delay effect on office employees. This served as indirect evidence that the introduction of staggered work hours or teleworking options could motivate office employees to travel during the off-peak hours.
Effects of distance (short vs. long distance)	The preference of short-distance travellers exhibited a more evenly distributed departure time compared to long-distance travellers. Moreover, the forecasting analysis shows a major shift of short distance car travellers towards off-	These findings suggest that policies aimed at managing congestion and optimising transportation resources may need to vary based on travel distance. For short-distance car travellers, strategies should focus on reducing travel time during off-peak hours and providing incentives to encourage off-peak travel, thus shifting demand away from peak periods. Additionally, promoting alternative transportation services available throughout the day can accommodate the varied departure time preferences of these travellers. For long-distance travellers, it may be beneficial to implement strategies to reduce the need for

Estimated parameters	Outcomes	Policy implication
	peak hours as the number of short-distance travellers increased.	longer car travel (such as park-and-ride facilities) (Marshall and Banister, 2000) or encourage teleworking and help spread out their departure times to reduce congestion during peak hours.
Effect of Income	Significant differences in correlations between departure times and durations among different income groups were found. Like the office employee, an increase in respondents from high-income households would result in higher demand around peak times.	High-income car commuters were more likely to choose peak hour travel. This group has higher affordability and is less likely to be price sensitive. This result complied with the findings from the study by Kockelman and Kalmanje (2005) in the context of Global South.
Sensitivity to travel time	Increase in travel time negatively impacted the utility (i.e., travellers preferred shorter travel times over longer ones) of car commuters.	To motivate car commuters to avoid peak hours and choose different times, it is important that travel during off-peak hours is consistently faster. If there is hardly any difference in travel times between peak and off-peak hours, it will become challenging to encourage people to explore other alternative times of the day as potential alternatives. In such cases, intervention to change departure time may not be an effective congestion management strategy.

3.7. Conclusions

This study presented a novel polynomial approximation of alternative specific constants (ASCs) to model departure time choice. The proposed functional form captured the interaction among different dimensions of time preference such as outbound, return, and duration. To the best of our knowledge, this was the first attempt to investigate the correlation between departure and duration within a departure time choice model framework. A joint departure time choice model (outbound and return) of car commuters was developed based on RP data from Dhaka, Bangladesh. The results indicated that the choices were significantly affected by the travel times and socio-demographic profile of the respondents. The proposed model reasonably agreed with the observed pattern.

The current study can be extended in several directions in the future. Firstly, the scope of this study was exclusively on car commuters. In the future research, it is crucial to apply our proposed modelling framework across various mode users and trip purposes. Such investigation will help examine the presence or absence of correlation in their time-of-day choice. Further, a similar structure can provide additional insights about non commute trips as well. It may be noted that the departure time choice for non-commute trips is more complex as there may be more flexibility associated with the choice of activity destination and mode. In such cases potentially warranting a joint model for departure time, destination, and mode. Our proposed polynomial functional form will serve as a foundational starting point for developing models that address the complexity of joint choice scenarios, accommodating multiple correlations across various activity and travel dimensions. Secondly, the lack of observed data post-implementation of strategies aimed at shifting time-of-day choices hinders the testing of the prediction accuracy of the proposed model. It is worth noting that, as of now, Dhaka has not implemented strategies such as congestion pricing, flexible working hours, time variant fares, etc. that could provide relevant data for such testing. In the future, after the implementation of any peak-spreading policies, the temporal preferences of car commuters can be compared with observed data, following the approach proposed by West et al. (2016) and Eliasson et al. (2013), to assess the accuracy and effectiveness of the proposed model's predictions. Thirdly, this study ignored the potential correlation between adjacent departure times to retain simplicity for practical implication. Future research can focus on estimating correlations of alternatives by using a more complex modelling framework, such as a cross-nested logit model or mixed MNL model, to account for the correlation among the alternatives. Finally, the current departure time choice model can be linked to a network assignment model specific to car users to evaluate potential route choice under dynamic traffic situation. Also, the research findings can be implemented in an agent-based simulation platform to take into account the interaction of car users with other mode users and test the impact of different peak spreading or congestion pricing policies.

Nevertheless, the proposed functional form overcame several issues related to the time preference model. As such, the proposed model (1) did not have the overfitting problem and gave more behaviourally realistic outputs, (2) reduced the computational cost by reducing the number of constants required to model time preference with the full set of constants, (3) addressed the issues associated with the correlation between departure time and the activity duration, (4) could accommodate multiple peaks without a priori assumption, and (5) had the potentiality to fit with both RP and SP data. The findings can be practically useful for devising peak-spreading policies in Dhaka — either as a stand-alone tool to test the impact of varied start times of offices in different locations or within an agent-based simulation tool to test the impact of different congestion pricing policies.

3.8. References

Abou-Zeid, M., Rossi, T.F. and Gardner, B. 2006. Modeling time-of-day choice in context of tour-and activity-based models. *Transportation research record.* **1981**(1), pp.42-49.

Afandizadeh Zargari, S. and Safari, F. 2020. Using clustering methods in multinomial logit model for departure time choice. *Journal of Advanced Transportation.* **2020**.

Anowar, S., Faghih-Imani, A., Miller, E.J. and Eluru, N. 2019. Regret minimization based joint econometric model of mode choice and departure time: a case study of university students in Toronto, Canada. *Transportmetrica A: Transport Science.* **15**(2), pp.1214-1246.

Arellana, J., Daly, A., Hess, S., de Dios Ortúzar, J. and Rizzi, L.I. 2012. Development of surveys for study of departure time choice: two-stage approach to efficient design. *Transportation Research Record.* **2303**(1), pp.9-18.

Arellana, J., de Dios Ortúzar, J. and Rizzi, L.I. 2013. Survey data to model time-of-day choice: methodology and findings. *Transport Survey Methods: Best Practice for Decision Making.* Emerald Group Publishing Limited.

Ashiru, O., Polak, J.W. and Noland, R.B. 2004. Utility of schedules: theoretical model of departure-time choice and activity-time allocation with application to individual activity schedules. *Transportation Research Record.* **1894**(1), pp.84-98.

Azhdar, R. and Nazemi, A. 2020. Modeling of incentive-based policies for demand management for the Tehran subway. *Travel Behaviour and Society.* **20**, pp.174-180.

Badiola, N., Raveau, S. and Galilea, P. 2019. Modelling preferences towards activities and their effect on departure time choices. *Transportation Research Part A: Policy and Practice.* **129**, pp.39-51.

Batur, İ. and Koç, M. 2017. Travel Demand Management (TDM) case study for social behavioral change towards sustainable urban transportation in Istanbul. *Cities.* **69**, pp.20-35.

Ben-Akiva, M. and Abou-Zeid, M. 2013. Methodological issues in modelling time-of-travel preferences. *Transportmetrica A: Transport Science.* **9**(9), pp.846-859.

Ben-Akiva, M. and Bierlaire, M. 1999. Discrete choice methods and their applications to short term travel decisions. *Handbook of transportation science.* Springer, pp.5-33.

Ben-Akiva, M. and Bierlaire, M. 2003. Discrete choice models with applications to departure time and route choice. *Handbook of transportation science.* Springer, pp.7-37.

Ben-Akiva, M., Litinas, N. and Tsunokawa, K. 1985. Continuous spatial choice: the continuous logit model and distributions of trips and urban densities. *Transportation Research Part A: General.* **19**(2), pp.119-154.

Bhat, C.R. 1998. Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological.* **32**(6), pp.361-371.

Bhat, C.R. and Steed, J.L. 2002. A continuous-time model of departure time choice for urban shopping trips. *Transportation Research Part B: Methodological.* **36**(3), pp.207-224.

Börjesson, M. 2008. Joint RP–SP data in a mixed logit analysis of trip timing decisions. *Transportation Research Part E: Logistics and Transportation Review.* **44**(6), pp.1025-1038.

Börjesson, M. 2009. Modelling the preference for scheduled and unexpected delays. *Journal of Choice Modelling.* **2**(1), pp.29-50.

Börjesson, M. and Kristoffersson, I. 2018. The Swedish congestion charges: Ten years on. *Transportation Research Part A: Policy and Practice.* **107**, pp.35-51.

Bwambale, A., Choudhury, C.F. and Hess, S. 2019. Modelling departure time choice using mobile phone data. *Transportation research part A: policy and practice.* **130**, pp.424-439.

Cervero, R. 1991. Congestion relief: the land use alternative. *Journal of Planning Education and Research.* **10**(2), pp.119-130.

Chaichannawatik, B., Kanitpong, K. and Limanond, T. 2019. Departure time choice (DTC) behavior for intercity travel during a long-holiday in Bangkok, Thailand. *Journal of Advanced Transportation.* **2019**.

Chin, A.T.H. 1990. Influences on commuter trip departure time decisions in Singapore. *Transportation Research Part A: General.* **24**(5), pp.321-333.

De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R. and Hofman, F. 2007. Uncertainty in traffic forecasts: literature review and new results for The Netherlands. *Transportation.* **34**(4), pp.375-395.

De Jong, G., Daly, A., Pieters, M., Vellay, C., Bradley, M. and Hofman, F. 2003. A model for time of day and mode choice using error components logit. *Transportation Research Part E: Logistics and Transportation Review.* **39**(3), pp.245-268.

De Palma, A., Hansen, P. and Labbé, M. 1990. Commuters' paths with penalties for early or late arrival time. *Transportation Science.* **24**(4), pp.276-286.

Ding, C., Mishra, S., Lin, Y. and Xie, B. 2015. Cross-nested joint model of travel mode and departure time choice for urban commuting trips: Case study in Maryland-Washington, DC region. *Journal of Urban Planning and Development.* **141**(4), article no: 04014036 [no pagination].

Dong, H. and Cirillo, C. 2020. Space-time dynamics: A modeling approach for commuting departure time on linked datasets. *Journal of Transport Geography.* **82**, p102548.

Eliasson, J., Börjesson, M., van Amelsfort, D., Brundell-Freij, K. and Engelson, L. 2013. Accuracy of congestion pricing forecasts. *Transportation Research Part A: Policy and Practice.* **52**, pp.34-46.

Geng, K., Wang, Y., Cherchi, E. and Guarda, P. 2023. Commuter departure time choice behavior under congestion charge: Analysis based on cumulative prospect theory. *Transportation research part A: policy and practice.* **168**, p103564.

Golshani, N., Shabanpour, R., Mohammadian, A., Auld, J. and Ley, H. 2019. Analysis of evacuation destination and departure time choices for no-notice emergency events. *Transportmetrica A: transport science.* **15**(2), pp.896-914.

Graham, D.J., Hörcher, D., Anderson, R.J. and Bansal, P. 2020. Quantifying the ex-post causal impact of differential pricing on commuter trip scheduling in Hong Kong. *Transportation Research Part A: Policy and Practice.* **141**, pp.16-34.

Guo, Y., Li, Y., Anastasopoulos, P.C., Peeta, S. and Lu, J. 2021. China's millennial car travelers' mode shift responses under congestion pricing and reward policies: A case study in Beijing. *Travel Behaviour and Society.* **23**, pp.86-99.

Habib, K.M.N., Day, N. and Miller, E.J. 2009. An investigation of commuting trip timing and mode choice in the Greater Toronto Area: Application of a joint discrete-continuous model. *Transportation Research Part A: Policy and Practice.* **43**(7), pp.639-653.

Habib, K.N. 2021. Departure Time Choice Modeling. In: Vickerman, R. ed. *International Encyclopedia of Transportation.* Oxford: Elsevier, pp.504-508.

Hendrickson, C. and Plank, E. 1984. The flexibility of departure times for work trips. *Transportation Research Part A: General.* **18**(1), pp.25-36.

Hensher, D.A. and Puckett, S.M. 2007. Congestion and variable user charging as an effective travel demand management instrument. *Transportation Research Part A: Policy and Practice.* **41**(7), pp.615-626.

Hess, S. and Palma, D. 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of choice modelling.* **32**, p100170.

Hess, S., Polak, J.W. and Bierlaire, M. 2005. Functional approximations to alternative-specific constants in time-period choice-modelling. In: *Transportation and Traffic Theory: Flow, Dynamics and Human Interaction, Proceedings of the 16th International Symposium on Transportation and Traffic Theory*: Elsevier Amsterdam, pp.545-564.

Hess, S., Polak, J.W., Daly, A. and Hyman, G. 2007. Flexible substitution patterns in models of mode and time of day choice: new evidence from the UK and the Netherlands. *Transportation.* **34**(2), pp.213-238.

Hightower, A., Ziedan, A., Crossland, C. and Brakewood, C. 2022. Current Practices and Potential Rider Benefits of Fare Capping Policies in the USA. *Transportation Research Record.* **2676**(10), pp.376-390.

Holyoak, N. 2008. *Departure time choice for the car-based commute*. thesis, Planning and transport research centre (PATREC).

Hossain, S., Hasnine, M.S. and Habib, K.N. 2021. A latent class joint mode and departure time choice model for the Greater Toronto and Hamilton Area. *Transportation.* **48**, pp.1217-1239.

Huan, N., Hess, S., Yao, E. and Liu, W. 2021. Time-dependent pricing strategies for metro lines considering peak avoidance behaviour of commuters. *Transportmetrica A: Transport Science.*

Jamal, S., Chowdhury, S. and Newbold, K.B. 2022. Transport preferences and dilemmas in the post-lockdown (COVID-19) period: Findings from a qualitative study of young commuters in Dhaka, Bangladesh. *Case studies on transport policy.* **10**(1), pp.406-416.

Khan, M., Choudhury, C.F. and Wang, J. 2011. Modeling preference for school bus service in Dhaka, Bangladesh: Stated preference approach. *Transportation research record.* **2239**(1), pp.64-73.

Kockelman, K.M. and Kalmanje, S. 2005. Credit-based congestion pricing: a policy proposal and the public's response. *Transportation Research Part A: Policy and Practice.* **39**(7-9), pp.671-690.

Koppelman, F.S., Coldren, G.M. and Parker, R.A. 2008. Schedule delay impacts on air-travel itinerary demand. *Transportation Research Part B: Methodological.* **42**(3), pp.263-273.

Lemp, J.D., Kockelman, K.M. and Damien, P. 2010. The continuous cross-nested logit model: Formulation and application for departure time choice. *Transportation Research Part B: Methodological.* **44**(5), pp.646-661.

Li, H., Li, X., Xu, X., Liu, J. and Ran, B. 2018. Modeling departure time choice of metro passengers with a smart corrected mixed logit model - A case study in Beijing. *Transport Policy.* **69**, pp.106-121.

Li, W., Cheng, D., Bian, R., Ishak, S. and Osman, O.A. 2018. Accounting for travel time reliability, trip purpose and departure time choice in an agent-based dynamic toll pricing approach. *IET Intelligent Transport Systems.* **12**(1), pp.58-65.

Li, Z. and Hensher, D.A. 2012. Congestion charging and car use: A review of stated preference and opinion studies and market monitoring evidence. *Transport Policy.* **20**, pp.47-61.

Marshall, S. and Banister, D. 2000. Travel reduction strategies: intentions and outcomes. *Transportation Research Part A: Policy and Practice.* **34**(5), pp.321-338.

McCafferty, D. and Hall, F.L. 1982. The use of multinomial logit analysis to model the choice of time to travel. *Economic Geography.* **58**(3), pp.236-246.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior.

Moya-Gómez, B. and García-Palomares, J.C. 2017. The impacts of congestion on automobile accessibility. What happens in large European cities? *Journal of Transport Geography.* **62**, pp.148-159.

Munch, E. and Proulhac, L. 2023. Is work hours' flexibility really a solution to morning peak period congestion? Comparative analysis between Paris and San Francisco. *Journal of Transport Geography.* **113**, p103712.

Newbery, D.M. 2005. Pricing and congestion: economic principles relevant to pricing roads. *Transport Economics.* Routledge, pp.265-287.

Noland, R.B. and Lem, L.L. 2002. A review of the evidence for induced travel and changes in transportation and environmental policy in the US and the UK. *Transportation Research Part D: Transport and Environment.* **7**(1), pp.1-26.

Palma, D., Enam, A., Hess, S., Calastri, C. and Crastes dit Sourd, R. 2021. Modelling multiple occurrences of activities during a day: an extension of the MDCEV model. *Transportmetrica B: Transport Dynamics.* **9**(1), pp.456-478.

Pan, J., Popa, I.S. and Borcea, C. 2016. Divert: A distributed vehicular traffic re-routing system for congestion avoidance. *IEEE Transactions on Mobile Computing.* **16**(1), pp.58-72.

Pucher, J., Peng, Z.r., Mittal, N., Zhu, Y. and Korattyswaroopam, N. 2007. Urban transport trends and policies in China and India: impacts of rapid economic growth. *Transport reviews.* **27**(4), pp.379-410.

Rahman, M., Murthy Gurumurthy, K. and Kockelman, K.M. 2021. Impact of Flextime on Departure Time Choice for Home-Based Commuting Trips in Austin, Texas. *Transportation Research Record.* p03611981211035757.

Saleh, W. and Farrell, S. 2005. Implications of congestion charging for departure time choice: Work and non-work schedule flexibility. *Transportation Research Part A: Policy and Practice.* **39**(7-9), pp.773-791.

Sasic, A. and Habib, K.N. 2013. Modelling departure time choices by a Heteroskedastic Generalized Logit (Het-GenL) model: An investigation on home-based commuting trips in the Greater Toronto and Hamilton Area (GTHA). *Transportation Research Part A: Policy and Practice.* **50**, pp.15-32.

Small, K.A. 1982. The scheduling of consumer activities: work trips. *The American Economic Review.* **72**(3), pp.467-479.

Thorhauge, M., Cherchi, E., Walker, J.L. and Rich, J. 2019. The role of intention as mediator between latent effects and behavior: application of a hybrid choice model to study departure time choices. *Transportation.* **46**(4), pp.1421-1445.

Thorhauge, M., Haustein, S. and Cherchi, E. 2016. Accounting for the Theory of Planned Behaviour in departure time choice. *Transportation Research Part F: Traffic Psychology and Behaviour.* **38**, pp.94-105.

Thorhauge, M., Swait, J. and Cherchi, E. 2020. The habit-driven life: accounting for inertia in departure time choices for commuting trips. *Transportation Research Part A: Policy and Practice.* **133**, pp.272-289.

Thorhauge, M., Vij, A. and Cherchi, E. 2021. Heterogeneity in departure time preferences, flexibility and schedule constraints. *Transportation.* **48**(4), pp.1865-1893.

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

TYPSA. 2019. *Feasibility study and preliminary design for construction of Dhaka subway.* Dhaka: TYPSA.

van der Loop, H., Haaijer, R. and Willigers, J. 2019. The impact of various forms of flexible working on mobility and congestion estimated empirically. *Autonomous vehicles and future mobility.* Elsevier, pp.125-139.

Vickrey, W.S. 1969. Congestion theory and transport investment. *The American Economic Review.* **59**(2), pp.251-260.

Wang, J.J. 1996. Timing utility of daily activities and its impact on travel. *Transportation Research Part A: Policy and Practice.* **30**(3), pp.189-206.

West, J., Börjesson, M. and Engelson, L. 2016. Accuracy of the Gothenburg congestion charges forecast. *Transportation Research Part A: Policy and Practice.* **94**, pp.266-277.

Yang, F., Jin, X. and Liu, R. 2008. Tour-Based Time-of-Day Choices for Weekend Nonwork Activities. *Transportation research record.* **2054**(1), pp.37-45.

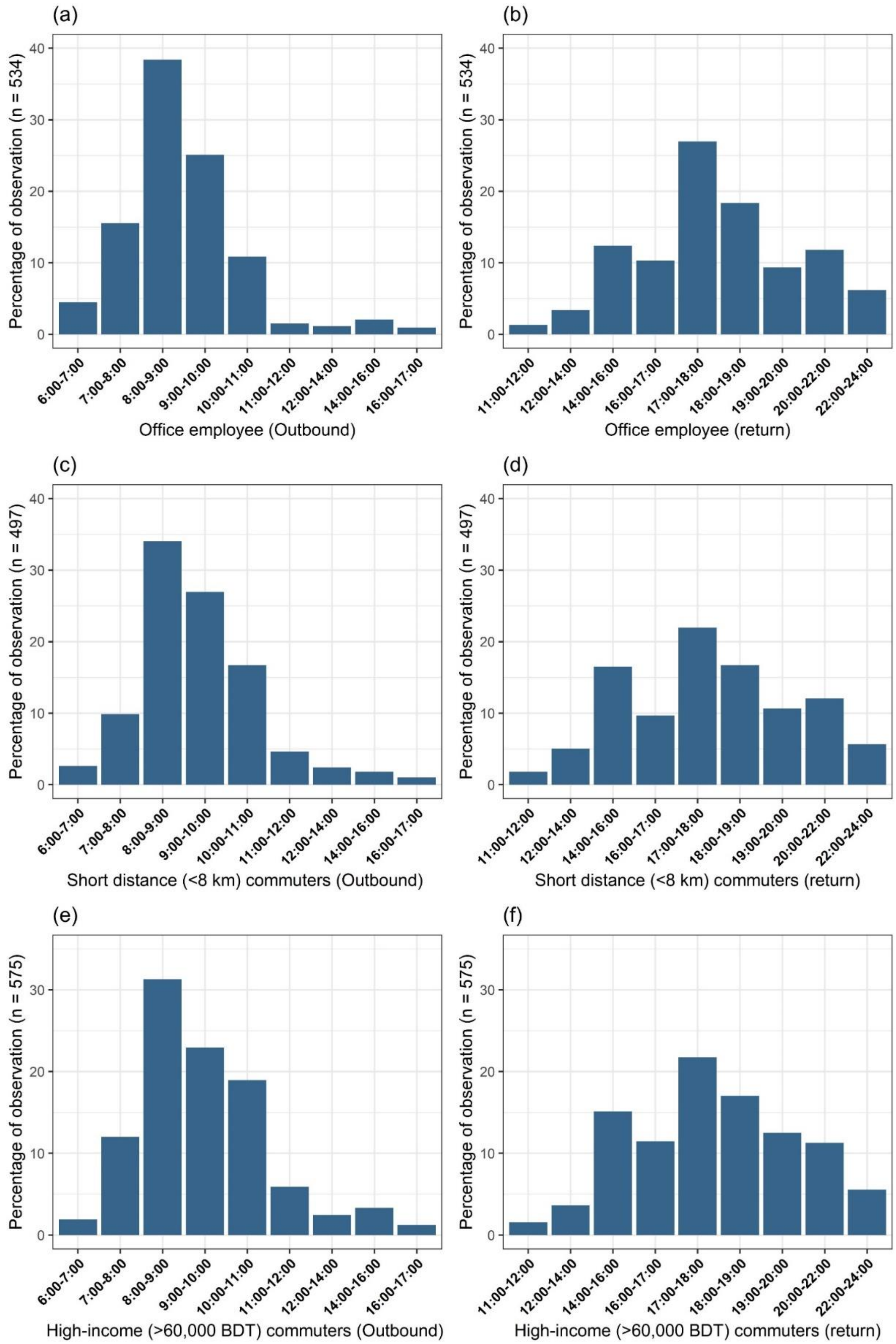
Yang, G. and Liu, X. 2018. A commuter departure-time model based on cumulative prospect theory. *Mathematical Methods of Operations Research.* **87**(2), pp.285-307.

Zannat, K.E., Choudhury, C.F. and Hess, S. 2022. Modeling departure time choice of car commuters in Dhaka, Bangladesh. *Transportation research record.* **2676**(2), pp.247-262.

Zhou, F., Li, C., Huang, Z., Xu, R. and Fan, W.D. 2020. Fare incentive strategies for managing peak-hour congestion in urban rail transit networks. *Transportmetrica A: Transport Science.* pp.1-37.

Zhu, Z., Chen, X., Xiong, C. and Zhang, L. 2018. A mixed Bayesian network for two-dimensional decision modeling of departure time and mode choice. *Transportation.* **45**(5), pp.1499-1522.

Appendix A



Supplementary Figure 3-1 Observed departure time choice of different socio-demographic groups.

Supplementary Table 3-1 Forecasting scenario results.

			Everyone behaves as if they were office employee		Everyone behaves as if they belong to high income households		Everyone behaves as if they were short distance traveller	
Alternatives	**Percentage**	**Base prediction**	**Forecast**	**Change**	**Forecast**	**Change**	**Forecast**	**Change**
6am_7am_11am_12pm	0.004	0.002	0.002	-20.92	0.001	**-58.23**	0.003	28.38
6am_7am_12pm_14pm	0.005	0.004	0.004	-2.71	0.002	-46.27	0.005	40.03
6am_7am_14pm_16pm	0.006	0.006	0.007	9.52	0.004	-35.80	0.009	45.36
6am_7am_16pm_17pm	0.003	0.006	0.007	12.21	0.004	-33.84	0.009	47.31
6am_7am_17pm_18pm	0.006	0.006	0.008	23.81	0.004	-40.77	0.009	48.09
6am_7am_18pm_19pm	0.002	0.006	0.007	23.87	0.003	-42.25	0.009	52.08
6am_7am_19pm_20pm	0.001	0.005	0.006	9.94	0.003	-40.07	0.008	54.75
6am_7am_20pm_22pm	0.009	0.006	0.006	8.85	0.003	-50.36	0.009	60.19
6am_7am_22pm_24pm	0.013	0.011	0.012	11.90	0.003	**-69.43**	0.018	61.32
7am_8am_11am_12pm	0.008	0.004	0.004	-19.86	0.003	-26.68	0.004	-18.11
7am_8am_12pm_14pm	0.012	0.009	0.009	0.42	0.008	-11.86	0.009	-1.09
7am_8am_14pm_16pm	0.018	0.017	0.019	13.69	0.017	0.12	0.019	10.68
7am_8am_16pm_17pm	0.011	0.018	0.021	16.13	0.019	3.32	0.021	13.96
7am_8am_17pm_18pm	0.027	0.018	0.023	26.51	0.017	-4.63	0.021	14.51
7am_8am_18pm_19pm	0.013	0.016	0.020	25.19	0.015	-7.80	0.019	15.50
7am_8am_19pm_20pm	0.014	0.014	0.016	9.90	0.014	-5.01	0.017	16.89
7am_8am_20pm_22pm	0.018	0.012	0.013	4.99	0.010	-16.07	0.015	21.18
7am_8am_22pm_24pm	0.008	0.014	0.014	0.33	0.008	-39.95	0.018	30.73
8am_9am_11am_12pm	0.003	0.006	0.004	-29.06	0.005	-13.67	0.003	-48.11
8am_9am_12pm_14pm	0.006	0.015	0.014	-3.91	0.014	-1.94	0.010	-29.05
8am_9am_14pm_16pm	0.029	0.032	0.036	14.44	0.034	6.09	0.028	-12.39
8am_9am_16pm_17pm	0.036	0.037	0.044	19.40	0.039	6.69	0.034	-7.05
8am_9am_17pm_18pm	0.084	0.075	0.104	**39.62**	0.078	4.32	0.070	-5.56
8am_9am_18pm_19pm	0.068	0.074	0.103	**39.58**	0.074	-0.26	0.070	-5.04
8am_9am_19pm_20pm	0.035	0.029	0.034	15.09	0.027	-6.88	0.028	-4.69
8am_9am_20pm_22pm	0.029	0.023	0.025	9.55	0.018	-20.56	0.023	-2.71
8am_9am_22pm_24pm	0.017	0.023	0.023	1.91	0.012	-47.30	0.024	5.48
9am_10am_11am_12pm	0.000	0.004	0.002	-51.58	0.004	-5.03	0.001	**-65.68**
9am_10am_12pm_14pm	0.007	0.011	0.008	-24.78	0.011	6.21	0.006	-45.06
9am_10am_14pm_16pm	0.032	0.026	0.026	-2.07	0.030	13.99	0.020	-22.75
9am_10am_16pm_17pm	0.027	0.033	0.034	4.79	0.038	14.98	0.028	-13.72
9am_10am_17pm_18pm	0.062	0.066	0.084	**27.94**	0.074	12.55	0.059	-10.65
9am_10am_18pm_19pm	0.031	0.032	0.034	3.83	0.035	7.60	0.029	-9.22
9am_10am_19pm_20pm	0.023	0.028	0.028	0.24	0.028	1.43	0.025	-8.73
9am_10am_20pm_22pm	0.031	0.021	0.019	-7.77	0.018	-12.08	0.019	-7.95
9am_10am_22pm_24pm	0.017	0.017	0.014	-20.93	0.010	-39.31	0.017	-2.82
10am_11am_11am_12pm	0.000	0.002	0.001	-74.22	0.002	-1.72	0.001	**-75.19**

			Everyone behaves as if they were office employee		Everyone behaves as if they belong to high income households		Everyone behaves as if they were short distance traveller	
10am_11am_12pm_14pm	0.011	0.006	0.003	-49.77	0.007	12.72	0.003	-53.02
10am_11am_14pm_16pm	0.029	0.015	0.010	-31.98	0.019	**24.53**	0.012	-24.15
10am_11am_16pm_17pm	0.017	0.021	0.016	-25.60	0.027	**29.26**	0.019	-10.27
10am_11am_17pm_18pm	0.025	0.022	0.019	-13.54	0.027	23.28	0.021	-4.73
10am_11am_18pm_19pm	0.026	0.021	0.018	-14.54	0.025	19.31	0.020	-1.52
10am_11am_19pm_20pm	0.016	0.020	0.013	-33.27	0.024	22.22	0.020	0.12
10am_11am_20pm_22pm	0.026	0.017	0.013	-27.00	0.020	13.72	0.017	0.92
10am_11am_22pm_24pm	0.015	0.011	0.004	-60.26	0.010	-6.15	0.011	2.51
11am_12pm_11am_12pm	0.001	0.001	0.000	**-83.37**	0.001	-7.85	0.000	**-72.01**
11am_12pm_12pm_14pm	0.004	0.002	0.001	-72.65	0.002	7.21	0.001	-59.76
11am_12pm_14pm_16pm	0.015	0.007	0.003	-52.77	0.008	20.31	0.005	-23.26
11am_12pm_16pm_17pm	0.003	0.010	0.006	-44.93	0.013	**25.61**	0.010	-2.80
11am_12pm_17pm_18pm	0.005	0.011	0.007	-33.38	0.013	20.00	0.012	6.28
11am_12pm_18pm_19pm	0.007	0.011	0.007	-33.23	0.013	16.35	0.012	12.52
11am_12pm_19pm_20pm	0.008	0.012	0.006	-49.54	0.014	19.42	0.013	15.97
11am_12pm_20pm_22pm	0.003	0.009	0.004	-59.12	0.010	10.52	0.011	17.47
11am_12pm_22pm_24pm	0.002	0.007	0.002	-73.50	0.006	-10.27	0.008	15.79
12pm_14pm_12pm_14pm	0.001	0.001	0.000	**-88.89**	0.001	8.46	0.000	-64.85
12pm_14pm_14pm_16pm	0.003	0.002	0.001	-73.21	0.002	9.47	0.001	-23.69
12pm_14pm_16pm_17pm	0.004	0.004	0.002	-64.06	0.005	15.08	0.004	5.37
12pm_14pm_17pm_18pm	0.003	0.005	0.002	-54.32	0.005	10.81	0.006	19.30
12pm_14pm_18pm_19pm	0.003	0.005	0.003	-52.07	0.006	7.62	0.007	29.92
12pm_14pm_19pm_20pm	0.004	0.006	0.002	-63.40	0.007	10.59	0.009	36.36
12pm_14pm_20pm_22pm	0.003	0.006	0.002	-70.63	0.006	1.16	0.008	40.11
12pm_14pm_22pm_24pm	0.002	0.005	0.001	**-82.27**	0.004	-20.69	0.007	34.43
14pm_16pm_14pm_16pm	0.000	0.000	0.000	-77.58	0.000	-0.54	0.000	-30.19
14pm_16pm_16pm_17pm	0.001	0.001	0.000	-65.82	0.001	6.27	0.001	-11.26
14pm_16pm_17pm_18pm	0.000	0.002	0.001	-54.30	0.002	4.23	0.002	12.32
14pm_16pm_18pm_19pm	0.004	0.002	0.001	-47.79	0.002	1.74	0.003	31.23
14pm_16pm_19pm_20pm	0.008	0.003	0.001	-54.68	0.003	3.57	0.004	44.13
14pm_16pm_20pm_22pm	0.007	0.004	0.001	-60.62	0.003	-5.48	0.006	53.77
14pm_16pm_22pm_24pm	0.009	0.004	0.001	-74.98	0.003	-27.12	0.006	47.76
16pm_17pm_16pm_17pm	0.000	0.001	0.000	-35.70	0.001	14.16	0.000	-58.12
16pm_17pm_17pm_18pm	0.000	0.001	0.001	-17.87	0.002	15.64	0.001	-31.46
16pm_17pm_18pm_19pm	0.002	0.002	0.002	-4.62	0.003	13.55	0.002	-4.59
16pm_17pm_19pm_20pm	0.000	0.004	0.004	-5.25	0.004	14.27	0.005	18.67
16pm_17pm_20pm_22pm	0.008	0.005	0.005	-7.92	0.006	7.25	0.007	41.38
16pm_17pm_22pm_24pm	0.003	0.006	0.004	-27.98	0.005	-11.89	0.009	45.70

Chapter 4 Joint modelling of activity type, start time and duration using a bounded multiple discrete-continuous framework: a case study of Dhaka, Bangladesh

Abstract

In activity-travel behaviour research, appropriate treatment of time dimension (e.g., time use) is an important prerequisite to forecasting the travel demand accurately. Two specific time dimensions of activity participation are highlighted in the literature: activity time use, and activity start time. There is limited research jointly investigated activity participation behaviour, considering both the choice of activity start time and activity duration. In this study, we attempted to investigate how individuals make decisions regarding activity type, start time and time use choice while facing constraint on their time allocation across different times of the day, particularly focusing on a developing country. A 24-h travel diary survey data from Dhaka, Bangladesh was used. Using the multiple discrete-continuous extreme value (MDCEV) frameworks, we estimated the activity type and duration model while accounting for utility differences based on activity start time. Additionally, we accommodated time budget constraints based on activity start time (morning or afternoon) using the bounded MDCEV model. Results showed that for the selected context, mandatory activities (e.g., work, education) had much greater utility differences based on activity start time than discretionary activities (e.g., leisure and social). Depending on the activity's starting time, utility and satiation associated with choice alternatives also varied across different socio-demographic groups (male, unemployed, etc.). Moreover, the application of bounded MDCEV ensured that no individual was predicted with time allocations higher than the imposed upper bound. From a policy perspective, the proposed modelling of activity type, start time and duration provides rich situational information about activity and travel patterns which can be used for activity timing, and agent-based simulation models.

Keywords: MDCEV, Bounded MDCEV, Activity Start Time, Activity Modelling, Travel Behaviour, Dhaka, Bangladesh

4.1. Introduction

Travel demand is a derived demand. Researchers have adopted different approaches, such as trip-based, tour-based, and activity-based frameworks, to understand travel behaviour. Over the past two decades, there has been a progressive transition from trip-based techniques (via tour-based approaches) to activity-based approaches in travel behavioural research (Bhat and Koppelman, 1999; Bowman and Ben-Akiva, 2001; Timmermans, 2005; Habib, K.M.N., 2011). This shift reflects a growing recognition of the significance of accounting for different activities and associated dimensions (e.g., activity type, duration, frequency, location etc.) that travellers consider while planning their journeys (Bhat and Koppelman, 1993; Bhat, 1996; Pawlak et al., 2017; Khaddar et al., 2022).

In activity-travel behaviour analysis, appropriate treatment of time dimension (e.g., time use, opening and closing time of activity location) is considered to be an important prerequisite for scheduling and forecasting the travel demand accurately (Pinjari and Bhat, 2010b). The literature highlights two specific time dimensions of activity participation: activity time use, and activity start time. There is limited research jointly investigating activity participation behaviour, considering both the choice of activity start time and activity duration (Habib, K.M. et al., 2008; Habib, K.M.N., 2012; Golshani et al., 2018). Bhat (2005) and Eluru et al. (2010) used the start time dimension (i.e., day of the week, time of the day) as an explanatory variable which interacted with the utility. However, such an approach raises the discourse on whether activity start time affects activity type and duration choice (given that an earlier start time provides a longer timeframe to fulfil a task) or if the activity type and duration choice affect the activity start time (due to the prolonged duration requirement or schedule constraint of a specific activity prompting the selection of an earlier time slot within the allocated time budget). A more reasonable approach to navigating this theoretical discourse involves delving into the intricate interplay among activity type, duration, and start time (Golshani et al., 2018). Additionally, the utility of activity participation among different market segments is influenced by various factors such as types of activities, associated flexibility (where only duration matters, not the starting time), and the opening and closing hours of activity locations (Habib, K.M.N. and Carrasco, 2011; Habib, K.M.N., 2012; Calastri et al., 2017; Varghese and Jana, 2019; Khaddar et al., 2022; Khan et al., 2022; Wu and Hong, 2022; Zhang and Yao, 2022) that may vary at different times of the day. Respondents may face constraints in accessing certain activities (Liu et al., 2017), due to fixed opening and closing times of locations. Mandated activity schedules, late penalties, and individual preferences can also impact the optimal activity start time. Ignoring temporal dependence and maximum time allocations for activities may lead to overestimating or underestimating activity demand for different market shares.

Further, the consideration of activity timing in activity modelling is important, particularly within the context of the Global South where the choice of start time for an activity may impose specific flexibilities or constraints. For instance, in many countries within the Global South, engaging in evening private practices is a prevalent culture for professionals like doctors, lawyers, and academics. In such cases, morning practices may be associated with institutional obligations, while evening practices provide a certain degree of flexibility to the decision maker. Therefore, in the transportation planning process, it is important to consider activity start time alongside other dimensions, such as activity type and duration. This holistic approach is crucial for understanding how utilities are accrued from (priorities are assigned to) specific activities, ultimately aiding in the forecasting of activity and travel demand. Furthermore, in the demand forecasting model, utilities are assigned to each activity based on activity type and duration (Axhausen et al., 2016; Balać et al., 2018). However, different socio-demographic groups may reveal distinct priorities depending on different activity start time. Therefore, demand prediction simulations should incorporate variations in activity preferences based on start times. This is crucial because travel demand impact on infrastructure differs for activities occurring in the peak compared to off-peak, emphasising the need to account for temporal variations in demand.

Pinjari and Bhat (2010b) also highlighted the importance of activity start time in understanding the related decisions (i.e., the when-dimension) of individual activity participation. They used multiple discrete-continuous nested extreme value model (MDCNEV) and applied it to analyse non-workers' out-of-home discretionary activity time-use and activity start time decisions on weekdays. Within different nesting structures of the model, they accommodated different combinations of discretionary activities and start time while capturing the correlation among alternatives. However, they ignore the fact that the allocation-budget for activities performed in the morning and in the evening cannot be the same[20]. In this context, this study goes further than Pinjari and Bhat (2010b) by not only accounting for utility differences depending on activity start time, but also by explicitly accommodating time of day specific time budget constraints. To the best of our knowledge, no prior studies investigated how individuals make decisions regarding activity type, start time and time use choice while facing limitations on their time allocation across different times of the day, particularly focusing on a developing country.

Additionally, in joint modelling, the majority of studies have focused on specific types of activities (either mandatory or discretionary) within a day. For instance, Golshani et al. (2018) and Pinjari and Bhat (2010b) applied their models solely to discretionary activities. For an accurate prediction of daily activities, it is essential to incorporate both mandatory and discretionary activities in joint modelling. Focusing solely on discretionary or mandatory activities in activity modelling may lead to erroneous predictions of the activity demand within the 24-hour budget constraint. In this study, we seek to fill this gap by jointly modelling activity type, start time and duration decision for both mandatory and discretionary activities selecting Dhaka, Bangladesh as a case study. Specifically, we attempt to answer the following two research questions:

1. How does the intricate interplay among the choices of activity type, duration, and start time influence the activity decision pertaining to both mandatory and discretionary activities?
2. In what manner does this interplay exhibit variations across different socio-demographic groups?

This article makes three key contributions. First, it offers a joint modelling framework to accommodate activity type, start time and duration while ensuring the limit for individual activity and total budget constraints. Second, it evaluates the significance of the modelling of both mandatory and discretionary activity types. Finally, it focuses on Dhaka as a case study, a challenging context to develop a robust activity type and duration choice model. The rest of the paper is organised as follows. In section 4.2 a short overview of the survey and data collection is given along with the sample statistics. In section 4.3, a

[20] This is not to say that there are different time-budgets at play. While Pinjari and Bhat (2010b) considered a 24-hour budget, they did not recognise that an activity occurring in a specific time-of-day period (say, evening) should not have the entire 24-hour budget available for allocation.

description of the modelling framework is given which is followed by the model estimation. The results from the model estimation are then presented with a critical discussion in section 4.4.

4.2. Data

The empirical data is drawn from the travel diary survey from Dhaka Metropolitan Region (RAJUK area), Bangladesh. The survey was administered by TYPSA[21] (https://www.typsa.com/en/) as part of the Dhaka Subway Project. The data was collected from Monday to Saturday between 28th February 2019 to 4th May 2019. During the surveys, each household member was asked about trips made during the previous working day (from Sunday to Thursday). The survey elicited information about household and individual sociodemographic (e.g., age, gender, education, occupation, household income, vehicle ownership), trip characteristics (e.g., departure time, access mode, main mode, travel time, trip purpose), location information (e.g., address of origin and destination). Trips of children under 6 years old were not recorded. Also, trips of less than 10 minutes of walking were not recorded during this survey. The survey was well planned to avoid trips made on weekends[22], public holidays, strike days, election days, and major events (like festivals). The details about the survey can be found in (Zannat et al., 2022). From this source, 17,291 (total number of respondents 70,768) households were selected for this study.

4.2.1. Survey and sample statistics

After cleaning the data for missing (and/or incorrect information), the final sample had information from 70,756 respondents (that corresponds to 17,291 households). The socio-demographic characteristics of the respondents are reported in Table 4-1. As observed, the majority of the respondents were aged between 25 to 60 (approximately 50%) i.e., the working-age group. The remaining age groups were appropriately represented by the sample while comparing with the census population distribution (Table 4-1). By encompassing a range of age groups (including working and non-working age), the selected sample would reflect the varied preferences and behaviours that individuals of different ages bring to their daily activities. Notably, around 6% of the population in Dhaka owns a car (Flavia and Choudhury, 2019), which is well reflected in the selected sample (approximately 5% of households own at least 1 car and less than 10% of households own at least one motorcycle). Sample statistics showed that about 35% of respondents were employed and, approximately 30% were unemployed (e.g., retired, housewife, searching for a job). The unemployed respondents were slightly overrepresented since in Dhaka approximately 23% of the total population was unemployed in 2013 (WPR, 2021). Also, from the income distribution, it can be observed that the proportion of lower-income households (i.e., with monthly household income below 20,000 BDT, which is approximately 12% of the total sample) were significantly below the national figures (approximately 40% of the overall population distribution). However, this is expected as well, since household income, in general, would be higher in the capital city (owing to better employment opportunities) as compared to other cities or rural areas in Bangladesh.[23]Additionally, the respondents' levels of education ranged from illiterate (9.86%) to highly educated (15.09%) with a bachelor's or master's degree from a university.

[21] TYPSA is a group of independent engineering, design, and consulting firms.
[22] In contrast to the Global North, Friday and Saturday are considered as weekends in Bangladesh.
[23] Information on the income distribution of the population of Dhaka was not available.

Table 4-1 Sample statistics *(at the individual level)* and population distribution.

Socio-demographic variables	Categories	% of respondents	% of total population
Age	<18	30.93	57.32*
	18-25	13.17	
	25-40	27.71	39.66*
	40-60	21.93	
	>=60	6.26	3.02*
Gender	Male	51.65	54.12*
	Female	48.35	45.88*
Household size	<=5	89.5	
	>5	10.5	
Car Ownership	No car	95.41	94**
	1 car	4.01	6**
	1+ car	0.58	
Motorcycle ownership	Not having motorcycle	91.74	
	1 motorcycle	7.85	
	1+ motorcycle	0.41	
Occupation	Student	27.17	
	Office-employee	17.99	
	Self-employed personnel	14.99	
	Housewife	25.23	
	Retired	2.79	
	Unemployed	1.79	
	Other	10.04	
Education	Primary	22.37	
	Up-to college degree	50.44	
	University degree	15.09	
	Other (vocational/madrasa)	2.24	
	Illiterate	9.86	
Income (in BDT)	<10,000	0.91	40***
	10,000-20,000	11.61	
	20,000-30,000	25.62	50***
	30,000-40,000	25.86	
	40,000-60,000	18.96	
	>60,000	17.04	10***

* The age and gender distribution for Dhaka city in extracted from WorldPop (www.worldpop.org). The population grid is selected for the year of 2019 to check the representativeness of our sample with the population distribution.

**The car ownership scenario in Dhaka city collected from the recent car ownership study focusing on Dhaka city by Flavia and Choudhury (2019)

*** The income distribution is at the country level collected from Census Data Bangladesh, 2011 (BBS, 2011; Rahman, 2016)

In the 24-h diary, respondents provided information about 10 different trip purposes viz. a) staying at home, b) work, c) work related activities, d) education (e.g., school/university/tuition), e) shopping, f) medical, g) personal issues (going to the bank, post office, etc.), h) leisure, i) accompaniment, and j) other discretionary activities. Work and work-related activities were aggregated into a single "work" category; medical and personal issues were aggregated as "personal activities", and leisure and other discretionary activities were consolidated into the category of "other activities". Additionally, all travel time designated for participating in activities was aggregated into a combined "travel" category. During the data screening, we removed activities that were not bound within a 24-h diary (e.g., activity started at night and ended the next day). Table 4-2 shows the summary of time use across various activity purposes in our sample. As observed, the average work duration in our sample is 8.86 hours, which represents a typical working day in Dhaka. Also, the proportion of respondents involved in discretionary activities (e.g., shopping, personal issues, accompanying and other social and leisure activities) was lower than participation in mandatory activities (work and education). Similarly, compared to mandatory activities, the average activity duration for non-essential activities was lower. It is evident from Table 4-2 that, approximately 50% of people travelled to perform at least one out-of-home activity and the others remained at home. This is because travel included the travel time required to reach a destination. In other words, if someone is not travelling that means that person is not performing an out-of-home activity. Eventually, the average trip frequency reflected that the majority of the respondents who participated in an out-of-home activity performed that activity once only except the travel (as it includes both outbound and returns).

Table 4-2 Average duration and frequency of activities (if participated) in the sample.

Activity type	**% of sample who engage in the activity**	**Average duration (in hour)**	**Average activity frequency**
Home	100.00	19.82	NA
Work	26.18	8.86	1.15
School	13.50	5.11	1.02
Shop	2.99	2.63	1.01
Personal issues	7.17	3.90	1.04
Accompany/Drop-off	1.71	3.19	1.23
Leisure and social	0.69	3.90	1.02
Travel	49.46	1.47	2.22

The additional activity dimension considered in this study was the activity starting time. The distribution of activity starting time among the selected sample is summarised in Table 4-3. It should be noted that Table 4-3 includes summary statistics of the first episode in case of both single and multiple occurrences of activities. Most of the respondents started their activities (both mandatory and discretionary) before noon. Among the activities started in the afternoon, the majority of them were discretionary activities. However, the scenario was the opposite for the second episode which means most of the activity starting time was in the afternoon, though only 4.6% of respondents were involved in multiple occurrences of activities (Table 4-4). While the initial starting time for leisure and social activities in the second episode may appear unconventional, the statistics appeared reasonable in the case of Dhaka. For instance, engaging in activities like a morning walk, attending a yoga class, or going to the gym (potential leisure activities) after dropping a child off at school aligned with the plausible routines of a housewife mother.

Table 4-3 Distribution of activity starting time of first episode (approximately 50% people of the total respondent).

Starting time	Work	School	Shop	Personal	Accompany	Leisure and social
Before 7:00	14.78	18.67	2.46	3.76	33.25	7.06
7:00 - 10:00	71.69	65.83	26.07	30.85	49.63	17.84
10:00 - 12:00	8.57	13.10	15.73	18.50	11.33	12.27
12:00 - 16:00	3.93	1.98	30.80	26.30	4.22	40.15
16:00 - 19:00	0.92	0.40	21.26	16.97	1.32	18.96
19:00 - 22:00	0.11	0.03	3.68	3.52	0.25	3.72
after 22:00	0.01	0.00	0.00	0.10	0.00	0.00
Sum	100	100	100	100	100	100

Table 4-4 Distribution of activity starting time of second episode (approximately 4.6% people of the total respondent)

Starting time	Work	School	Shop	Personal	Accompany	Leisure and social
Before 7:00	0.24	0.00	0.00	0.00	0.00	0.00
7:00 - 10:00	4.13	2.78	10.00	3.55	14.04	14.29
10:00 - 12:00	13.33	9.44	3.33	13.61	43.84	14.29
12:00 - 16:00	64.10	62.22	53.33	52.07	29.79	42.86
16:00 - 19:00	15.93	23.89	23.33	26.63	9.25	28.57
19:00 - 22:00	2.16	1.67	10.00	4.14	3.08	0.00
after 22:00	0.12	0.00	0.00	0.00	0	0.00
Sum	100	100	100	100	100	100

4.3. Model Specification

4.3.1. Modelling Framework

4.3.1.1. Classical MDCEV model

The MDCEV model is derived based on the random utility maximisation theory, where decision makers are assumed to allocate resources so as to maximise the accrued utility (represented by $U(x)$) subject to a set of budget constraints. The advantage of the MDCEV model over the classical choice model is that it can capture the satiation effects of continuous consumption separately from the utility that determines the discrete decision. Also, the MDCEV framework relaxes the mutual exclusivity assumption of the classical discrete choice model, while at the same time still allowing for corner solutions (i.e., where one or more goods are not consumed).

The utility equation introduced by Bhat (2008) is given as follows:

$$U(x) = \underbrace{\frac{1}{\alpha_1}\psi_1 {x_1}^{\alpha_1}}_{\text{Sub-utility for outside goods}} + \underbrace{\sum_{k=2}^{K} \frac{\gamma_k}{\alpha_k}\psi_k \left(\left(\frac{x_k}{\gamma_k}+1\right)^{\alpha_k} - 1\right)}_{\text{Sub-utility for inside goods}} \qquad \text{(i)}$$

U(x) is quasi-concave and continuously differentiable with respect to the consumption quantity vector *x*. ψ_k is the baseline utility of good *k* i.e., the marginal utility of the good at the point of zero consumption. ψ_k is the function of observed characteristics of decision maker and good *k*. ψ_k can be further parameterised as:

$$\psi_k = e^{\delta_k + \beta_k * z_k + \varepsilon_k} \qquad \text{(ii)}$$

δ_k is constant for alternative *k* for the baseline utility, z_k is the vector of attribute associated with the alternative (activity type) and characteristics of the decision maker. ε_k is the unobserved characteristics that affect the baseline utility for good *k*.

α_k and γ_k are related to good *k*. The role of α_k is to reduce the marginal utility with increasing consumption of good *k*. If $\alpha_k = 1$ for all *k*, this represents the case of absence of satiation effect or constant marginal utility. γ_k is the translation parameter which allows for corner solution and is also associated with the level of satiation (larger value of γ_k for alternative *k* implies lower satiation i.e., a higher consumption when chosen). Since, α_k and γ_k both has influence on the consumption quantity of alternative *k*, it is difficult to differentiate the effect of α_k and γ_k parameter from the same model. To avail a more efficient forecasting algorithm, many studies used γ_k profile (with $\alpha_k \rightarrow 0$ for all alternatives) for model estimation (Pinjari and Bhat, 2010a; Calastri et al., 2017; Palma et al., 2021).

And γ_k is further parameterised as:

$$\gamma_k = \theta_k + \lambda_k * w_k \qquad \text{(iii)}$$

θ_k is constant for alternative *k* and λ_k corresponds to the parameter capturing the impact of sociodemographic attribute w_k.

Also, the choice of total consumption amount is subject to budget constraint (*total budget*, *B* = $\sum_{k=1}^{k} x_k p_k$, p is the vector of unit price). The likelihood that an individual chooses a specific vector of consumption amount $({x_1}^*, {x_2}^*, \ldots\ldots, {x_M}^*, 0, \ldots, 0)$ (assuming IID type I extreme value distribution for the stochastic component in the utility function):

$$L({x_1}^*, {x_2}^*, \ldots\ldots, {x_M}^*, 0, \ldots, 0) = \frac{1}{p_1}\frac{1}{\sigma^{M-1}}\left(\prod_{m=1}^{M} f_m\right)\left(\sum_{m=1}^{M}\frac{p_m}{f_m}\right)\left(\frac{\prod_{m=1}^{M} e^{\frac{V_m}{\sigma}}}{\left(\sum_{k=1}^{K} e^{\frac{V_k}{\sigma}}\right)^M}\right)(M-1)! \qquad \text{(iv)}$$

In the above expression,

$$V_k = \beta_k * z_k - \ln\left(\frac{x_k^*}{\gamma_k p_k} + 1\right) - \ln(p_k)\,;\; k \geq 2, \quad \text{(v)}$$

and, p_k is the unit price of good k, and α is fixed to 0.

$$f_m = \frac{1}{x_m{}^* + \gamma_m} \quad \text{(vi)}$$

Also, for the outside good,

$$V_1 = -\ln\,(x_1^*)\; - \ln(p_k) \quad \text{(vii)}$$

4.3.2. MDCEV with bounds

While non-negativity constraints ($x \geq 0$) and a binding budget constraint on allocations ($\sum_{m=1}^{M} x_m = total\ budget$) are taken into account in the optimisation of the classical MDCEV model, the bounded MDCEV model also imposes additional inequality constraints, such as upper bounds and/or lower bounds, on the optimisation. In reality, decision-makers encounter such constraints (upper bound and lower bound constraints) while accommodating multiple activities within a limited time period. For instance, a strict inequality in the distribution of time throughout the day may be a result of a rigid workplace schedule, a family duty, a deadline, or an individual's preference or ability.

For upper bounds:

$$x_i^* \leq x_i^{max} \text{ for } i\epsilon\ goods\ with\ upper\ bounds \quad \text{(viii)}$$

Here, x_i^{max} is exogenously determined upper bound on the consumption for alternative *i*.

However, the allocations should meet the binding budget constraint. The utility structure of the bounded MDCEV model will follow the same additively separable utility profile of classical MDCEV model (equation i). The baseline utility of good k (ψ_k) and translation parameter (γ_k) is parameterised following equation (ii) and equation (iii). Though for the outside good V_1 can be defined same as the classical MDCEV framework, however, in the bounded MDCEV, V_k can be defined following different equality and inequality constraint:

$$V_k = \beta_k * z_k - \ln(p_k) \qquad \text{if } x_k^* = 0$$

$$V_k = \beta_k * z_k - \ln\left(\frac{x_k^*}{\gamma_k} + 1\right) - \ln(p_k) \qquad \text{if } 0 < x_k^* \leq x_k^{max} \quad \text{(ix)}$$

Assuming the error terms to be IID Gumbel distributed, the likelihood of the MDC model with explicit upper bounds (as in Equation (viii)) can be derived to be an analytically tractable, closed form expression.

The details of the optimality conditions with upper bounds (from Equation (viii)) and the derivation of the resulting likelihood function can be found in Saxena et al. (2021).

4.3.3. Modelling Issues

In the utility specification, along with the activity type, we added a new dimension i.e., activity start time. We divided each activity type based on their different start time (for example, the work activity started in the morning and the work activity started in the afternoon) allowing differences in their utility and satiation. Figure 4-1 shows the classification of activities used in the modelling framework. Activities starting before noon had a maximum possible allocation of 24 hours and activities starting in the afternoon period had a maximum allocation of 12 hours. We fixed an upper bound of 12 hours for afternoon activities since the detailed description of the out-of-home activities was not specified in the travel diary survey. Though morning and afternoon activities had different upper bounds, the aggregated consumption was limited to 24 hours. Three model formulations were explored in this study — 1) a classical MDCEV model where activities were not classified depending on activity start time, 2) a classical MDCEV model where activities were classified based on start time without time-of-day specific budget constraints, and 3) a bounded MDCEV model by classifying and imposing bounds depending on their start time. The second and third models allowed for different utilities and satiation for activities starting in the morning and afternoon, but only the third model recognised that the maximum time allocation for afternoon activities was less than for those starting in the morning. All three models of course still imposed the 24-hour total budget limit. These three models, including only baseline constants and satiation parameters, were considered as the base model. They were compared with each other to highlight the importance of including activity start time dimensions and bounds on different activity types. To choose the model for interacting socio-demographic variables, the predicted time use from the base estimated models and observed time use were compared. The final specification was chosen from the model that best fit the observed time usage and offered logical forecasting possibilities.

Based on statistical significance (used LR test), several socio-demographic factors were gradually added to the base model. The model fit and behavioural consistency of the estimated parameters were taken into account for the final specification. All the socio-demographic variables were included in the utility equation as a dummy variable. For instance, the categories “less than 25”, “25-40”, “40-60”, and “beyond 60” were used to categorise age. We also tested the effect of gender and working status (as unemployed e.g., housewife, retired, and looking for a job) at the individual level. Additionally, having underaged children (0–5 years old), owning a car or motorcycle, having a household monthly income of at least 60,000 BDT, living in a TAZ with a higher wealth index than 1.5, and having a household of at least four people were all considered at the household level. Since vehicle ownership (either having a car or motorcycle) and income level were not highly correlated to each other (Spearman correlation value $r = 0.24$), both were used as covariates in this study. One of the reasons for having less correlation between income and vehicle ownership could be the use of ride-hailing services in Dhaka which may have increased both car and motorcycle ownership among the lower-middle-income people considering it as a source of income (Wadud, 2020). However, the level of education was highly correlated with the household monthly income. Therefore, we estimated the influence of income on activity preference to avoid the overfitting problem which may be caused by using them both in the same model. To identify the effect of residential location, we classified the TAZs based on the wealth index (a poverty indicator) provided by (Steele et al., 2017).

The estimation was done using GAUSS 22 licensed version using Maximum Likelihood estimation. During the model estimation, all the alternatives either starting in the morning or the afternoon were made available to all socio-demographic groups. This is because survey respondents were not asked whether or not their activities were time sensitive. Furthermore, it was unknown if the participants had a choice of activities during the weekday. Also, sample statistics showed the distribution of different market share performing activities at different times of the day (Supplementary Table 4-1). Retired people, housewives, and unemployed people all participated in work-related activities. Retired employees could work part-time after they retire, unemployed people could participate in specific work-related activities to find jobs, and housewives could be involved in specific cooperative organisations. To avoid misleading assumptions regarding the availability of activities for different market shares, we made all activities available for everyone at different times of the day. This approach allowed us to accommodate the choice of respondents who participated in specific activities more than once by beginning at various times throughout the day in our model.

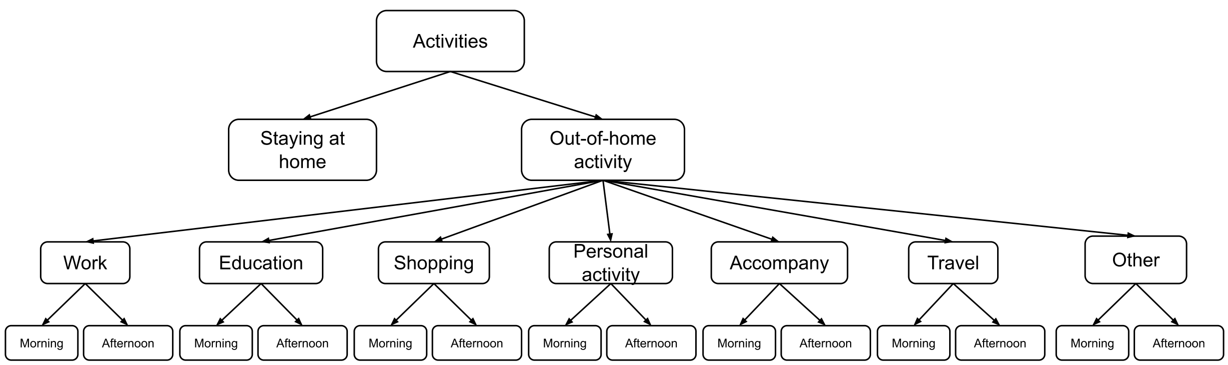


Figure 4-1 Choice set of activities.

4.4. Results and discussion

4.4.1. Base models

Table 4-5 shows the estimates of the three base models developed using classical MDCEV and bounded MDCEV models. Baseline utility constants in these models, or how frequently the sample chooses the activity, were consistent with discrete choices. The alternative specific constants (ASC) in the specification do not have much behavioural significance. However, the negative signs of all ASCs corresponding to inside goods represent that all individuals in our sample spent some non-zero time at home (the essential outside good). The addition of utility differences in the MDCEV model (classical MDCEV and bounded MDCEV) allowed us to determine what percentage of respondents chose to engage in out-of-home activities during the first and second half of the day. While considering certain activities at a specific time of the day, a notable difference in utility was observed for some activities. For instance, people were more likely to pick the morning for work, education, and accompanying, whereas they were more likely to choose the afternoon for personal and shopping activities. This could be because of the less flexible nature of morning activities such as work and education that cannot be performed in the afternoon due to the prevalent practice of conventional office-business hours (9:00 – 17:00) and school schedules. Moreover, accompanying someone is a type of shared responsibility (e.g., dropping off children at school or accompanying parents), more often convenient or required to perform in the morning due to synchronise schedule (office and school starting time) and meet mobility requirements. For example, it is a common practice in Bangladesh to use a car among different household members (Enam and Choudhury, 2011). However, the baseline preference for travelling, leisure and social activities was homogenous in the morning and afternoon compared to the other out-of-home activities.

Further, the co-efficient of the satiation parameter as shown in Table 4-5 suggested that levels of satiation varied depending on the activity type and the corresponding start time. The results from the MDCEV model (without time-varying utility) noted that respondents were likely to spend longer time on mandatory activities compared to discretionary activities. On the other hand, the timewise utility model (both classical MDCEV and bounded MDCEV) demonstrated that respondents were less rapidly satiated for the majority of the morning activities, such as work, education, personal activities, leisure, and accompanying activities. For other activities such as travelling, and shopping, the effects of satiation appeared to be uniform both in the morning and afternoon. It is important to note that compared to other activities, those showing a uniform satiation level throughout the day had a comparatively shorter time requirement (Table 4-2).

Besides the signs and magnitude of the estimated parameters, we also compared the prediction accuracy of the estimated models. Root mean square error (RMSE) was calculated to compare the difference between time use predicted by the estimated models and observation. The prediction of the classical MDCEV model was done following the algorithm proposed by Pinjari and Bhat (2010a). The bounded MDCEV prediction was done using the upper bound prediction algorithm proposed by Saxena et al. (2021). Table 4-6 shows the RMSE of the MDCEV model without the activity start time, the MDCEV model with activity start time (without activity and time specific budget constraint), and the bounded MDCEV model with activity start time (with activity and time specific budget constraint). Table 4-6 includes both time-wise activity specific RMSE and aggregated RMSE (combining morning and afternoon activity time use). The activity specific RMSE values from morning and afternoon activities exhibited relatively less error than the aggregated RMSE values. Overall, the RMSE values from three estimated models indicated that the bounded MDCEV model, accommodating both time-varying utility and time-of-day specific time budget constraints, showed relatively less error than the other model. However, while comparing the difference in RMSE values between classical MDCEV and bounded MDCEV (with activity start time) using t-test, the differences were not statistically significant. Nevertheless, for the final specification, we used bounded MDCEV with upper bounds on consumption to make the model logically consistent and ensure that no individual was predicted with time allocations higher than available time later in the day.

Table 4-5 Model summary of base Models.

MDCEV (without activity start time)		MDCEV (with activity start time)		
Parameter	**Estimates MDCEV**	**Parameter**	**Estimates MDCEV**	**Estimates Bounded MDCEV**
Baseline utility constants				
Work	-3.981***	Morning	-3.972***	-3.972***
		Afternoon	-5.651***	-5.656***
Education	-4.694***	Morning	-4.717***	-4.717***
		Afternoon	-7.117***	-7.119***
Shopping	-6.276***	Morning	-7.183***	-7.183***
		Afternoon	-6.714***	-6.715***
Personal	-5.380***	Morning	-6.075***	-6.075***
		Afternoon	-5.959***	-5.960***
Accompany	-6.847***	Morning	-6.904***	-6.905***
		Afternoon	-8.218***	-8.218***
Leisure	-7.775***	Morning	-8.663***	-8.662***
		Afternoon	-8.195***	-8.195***
Travel	-2.772***	Morning	-2.806***	-2.806***
		Afternoon	-2.667***	-2.667***
Baseline satiation constants				
Work	14.708***	Morning	12.061***	12.066***
		Afternoon	5.750***	5.694***
Education	6.604***	Morning	6.715***	6.715***
		Afternoon	4.729***	4.723***
Shopping	2.508***	Morning	2.416***	2.418***
		Afternoon	2.632***	2.628***
Personal	3.987***	Morning	5.329***	5.332***
		Afternoon	3.189***	3.181***
Accompany	3.156***	Morning	2.708***	2.707***
		Afternoon	1.712***	1.710***
Leisure	4.091***	Morning	4.589***	4.588***
		Afternoon	3.769***	3.772***
Travel	0.550***	Morning	0.253***	0.253***
		Afternoon	0.287***	0.287***
Goodness of fit				
N. parameters	14		28	28
Log-Likelihood (Start)	-615185.9193		-783453.7	-783302.2
Log-Likelihood (Final)	-303185.2146		-369266.7	-369098.0

*** All the values are significant at 95% confidence level, ** All the values are significant at 90% confidence level, * All the values are significant at 80% confidence level.

Table 4-6 Model comparison between MDCEV with and without time of the day specific budget constraint.

		Stay at home	**Work**	**Education**	**Shopping**	**Personal**	**Accompany**	**Leisure**	**Travel**	**Overall**
RMSE of MDCEV without activity start time		8.48	6.84	3.77	1.27	2.35	1.05	0.72	3.45	4.38
RMSE of MDCEV with activity start time	Morning	8.54	6.50	3.67	0.76	1.91	0.96	0.50	2.29	3.35
	Afternoon		2.35	0.97	0.96	1.56	0.46	0.58	2.64	
	Aggregated	8.54	6.86	3.79	1.22	2.46	1.07	0.76	3.61	4.30
RMSE of MDCEV with bounds for afternoon activity	Morning	8.41	6.51	3.66	0.76	1.86	0.92	0.50	2.32	3.16
	Afternoon		1.94	0.83	0.85	1.34	0.33	0.49	2.22	
	Aggregated	8.41	6.76	3.71	1.14	2.28	1.04	0.72	3.34	4.20

4.4.2. Bounded MDCEV model with sociodemographic factors

Results of the full specification of the bounded MDCEV model are summarised in Table 4-7 and Table 4-8. The final Log-Likelihood value of full-specification (-355914.71) demonstrates how the base model was improved by the addition of socio-demographic variables, strongly indicating that these effects offer unique insights into behaviour.

4.4.2.1. Baseline utility

The baseline preference constants and the effects of socio-demographic factors on the baseline utilities are shown in Table 4-7. Like the base bounded MDCEV model, the negative value of constants emphasised the higher utility for the base alternative — staying at home — over any other activities that took place outside the home. After the inclusion of socio-demographic variables in an additive manner with the baseline utility, the constant (denoted as δ) represented the base demographic group, serving as the reference category for comparative analysis. This base demographic group, encompassing attributes such as age >60, females, employed individuals, households without a vehicle, and those with an income less than or equal to 60,000 BDT, etc. established the benchmark against which other demographic groups were assessed. A positive coefficient associated with a specific socio-demographic variable implies a positive shift or preference for the activity choice within that demographic, relative to the reference group. Conversely, a negative coefficient indicates a greater preference for the activity within the reference group. This analytical approach provides nuanced insights into the comparative preferences and decision-making dynamics across diverse socio-demographic segments.

Individual sociodemographic

Activity start time and its association with the choice of alternative activities were more evident while considering the socio-demographic interaction with the baseline utility parameters. Respondents below 25 years old (approximately 45% of the total sample) were more likely to study compared to older adults (age >60 years). Also, the youngest age group had the higher utility from education while starting the activity in the morning. It is plausible that this age group under consideration was not a good fit for evening-time educational institutions, starting the education activity in the morning would be the most beneficial choice. Respondents, who preferred afternoon educational activities, might be pre-occupied in the morning inducing higher utility for educational activity in the afternoon. Further, the youngest age group was less likely to participate in leisure activities which also include social activities. This implies that during a working day, young people were less likely to be involved in discretionary activities compared to older people due to their higher propensity to engage in mandatory educational activities (e.g., going to school, tuition, etc.). The heightened commitments to educational activities and the decreased probability of engaging in out-of-home discretionary activities underscore the unique activity patterns and priorities observed within this age group on working days. These findings emphasised the influential role of educational commitments in shaping the discretionary activity engagement of younger individuals (Habib and Daisy, 2013).

For the respondents aged between 25 to 40 years old, they were more likely to choose work activities, personal activities, and accompanying someone. When it comes to selecting the obligatory tasks (from education to work), there was a substantial difference between decision-makers under the age of 25 and those between the ages of 25 and 40 (Table 4-7). Compared to older adults (over the 60s), respondents from 25 to 40 years old may benefit from other secondary activities like accompanying someone to a place or engaging in their personal activities. Within this age range, there was also notable variability in the time that activities began. They had the highest utility from the work if the activity was started before noon, compared to people "over the 60s". For secondary activities such as accompanying or performing personal activities, the highest utility was achieved while they were performing those activities in the morning.

A gender effect was also tested in the baseline utility. Gender effect appeared significant for personal activity and accompanying related activities. Male respondents were less likely to accompany both in the morning and afternoon. Besides, they were less likely to perform a personal activity in the morning compared to female respondents. This is plausible because Bangladesh is a male dominated country (Andaleeb and Vanneman Wolford, 2004), and it is reasonable that they were likely to invest their time in mandatory activities rather than personal activities.

Furthermore, employment status was also found to be associated with the likelihood of performing an activity at a different time of the day. The most popular activities among the unemployed people (e.g., housewives, retired, etc.) were shopping and leisure activities. These findings complemented the finding from the age and gender effects on baseline preference. Compared to the employed respondents, unemployed people were more likely to perform these discretionary activities in the morning. This group of people had the flexibility to conduct discretionary activities at the time of the day when employed people were likely to be engaged in mandatory activities.

Household sociodemographic

Like the individual factors (gender, employment status), household covariates were found to be sensitive to the activity start time and activity decision. For instance, the decision to do work and accompany other people was significantly affected by household factors such as the presence of underaged children (0–5 years) in the household. The results showed that people having underaged children in the household were less likely to work in the morning (Table 4-7). The increased household responsibilities and lack of reliance on daycare centres in Dhaka could be plausible reasons. Additionally, family members having young children were more likely to accompany in the morning and less likely to accompany in the evening. The added childcare responsibilities in the morning to accompany them for out-of-home activities (e.g., school) could be associated with the time preference for other activities.

Besides having underaged children in the household, the effect of household vehicle ownership was also tested to understand the significance of activity start time while making activity decisions. The results indicated that respondents from households having a personal vehicle (e.g., motorbike or car) were more likely to engage in personal and leisure activities during a typical working day compared to respondents who do not own a private vehicle. Respondents' preference to be involved in these discretionary activities in the afternoon reinforced the added possibilities of higher mobility of vehicle owners. In relation to income level, high-income people (those earning more than 60,000 BDT) had the highest utility for shopping in the afternoon during a working day, compared to middle- and low-income people. People from middle- and low-income groups would typically have the predisposition to maximise their benefit from work and other necessary activities related to their daily lives due to their greater financial strain. Therefore, it can be inferred that in Dhaka's case, both household income and vehicle ownership were related to a larger propensity to engage in discretionary activities in the afternoon.

The impact of home location on activity decisions was investigated along with household variables. To protect the respondents' privacy, home addresses were taken into account at the TAZ level. In Dhaka, the average wealth of TAZ and household income were not correlated due to the heterogeneous land use pattern. The final specification, therefore, considered the effects of both wealth of TAZ and household income. Contrary to individuals who reside in less wealthy areas, residents of wealthy TAZs were likely to accompany and participate in leisure and social activities. In Dhaka, wealthy TAZ refers to highly commercialised areas. The availability of indoor and outdoor social or recreational facilities was higher for people inhabiting in wealthy areas. Besides, respondents in those areas preferred to accompany more due to their feeling of insecurity associated with commercial areas (e.g., high traffic volume). Significant heterogeneity was also observed among the respondents in terms of choosing the activity type and corresponding starting time while considering the home location. Residents of wealthier areas were more likely to participate in out-of-home activities in the afternoon, compared to people from less wealthy areas. They were also likely to choose the afternoon for their discretionary activities. People living in wealthy TAZ would need to travel less compared to the other to reach the discretionary destinations as commercial areas were located adjacent to their home location. Therefore, they had the opportunity to engage both in leisure and social activities in the afternoon.

Table 4-7 Bounded MDCEV results utility parameters.

	Activity		**Estimate**	**Rob.t.rat.(0)**
Baseline utility constants	Work	Morning	-4.199	-228.716
		After noon	-5.863	-209.288
	Education	Morning	-7.519	-124.288
		After noon	-9.012	-81.112
	Shopping	Morning	-7.844	-108.806
		After noon	-7.298	-134.586
	Personal activity	Morning	-5.775	-141.134
		After noon	-5.888	-150.839
	Accompany	Morning	-7.094	-87.805
		After noon	-8.366	-51.799
	Leisure	Morning	-9.723	-51.167
		After noon	-9.449	-59.05
	Travel	Morning	-2.773	-176.046
		After noon	-2.630	-206.566
Age <25	Education	Morning	4.095	69.166
		After noon	2.871	24.642
	Leisure	Morning	-1.585	-8.025
		After noon	-0.796	-5.725
Age 25-40	Work	Morning	0.773	31.77
		After noon	0.576	15.623
	Personal	Morning	1.851	23.621
		After noon	1.698	10.884

	Accompany	Morning	0.391	7.895
		After noon	0.295	6.984
Gender = male	Personal activity	Morning	-0.770	-18.019
		After noon	-0.348	-8.737
	Accompany	Morning	-3.024	-26.396
		After noon	-2.667	-12.986
Occupation=Unemployed	Shopping	Morning	1.213	15.22
		After noon	0.826	13.957
	Leisure	Morning	1.857	10.481
		After noon	1.020	7.982
Having underaged children at the household	Work	Morning	-0.078	-3.755
	Accompany	Morning	0.150	2.281
		After noon	-0.184	-1.381
Vehicle ownership=”Yes”	Personal	Morning	0.073	1.255
		After noon	0.325	6.367
	Leisure	After noon	0.323	2.301
Income >60,000 BDT	Shopping	Morning	-0.107	-1.461
		After noon	0.324	5.592
Home location (WI>1.5)	Accompany	Morning	0.360	5.601
		After noon	0.588	4.617
	Leisure	Morning	0.733	4.62
		After noon	1.451	10.015

4.4.2.2. Translation parameters

Table 4-8 summarises the final model’s coefficient, explaining the different levels of satiation from various activities across different socio-demographic groups commencing at different times of the day (morning/afternoon). Like the baseline utility, θ_k represents the baseline satiation (i.e., satiation for the reference group) and λ_k signifies the impact of specific socio-demographic characteristics (i.e., the shift from the base value of θ_k). Here, reference categories used in the baseline utilities remained consistent during the estimation of the translation parameters. To estimate θ_k and λ_k , γ_k was further parameterised as $exp\ (\theta_k + \lambda_k * z_k)$. In Table 4-8, baseline satiation (γ_k) shows the value of $exp\ (\theta_k)$ and λ_k shows the shift from the log of the baseline γ_k. A positive shift in the translation parameter for alternative *k* indicates a reduced rate of satiation, suggesting a greater willingness to invest more time in activity *k*. On the other hand, negative values decrease the translation parameter, implying a faster satiation and a tendency for shorter time investment in activity *k*.

Impact Individual sociodemographic on satiation

The difference in satiation associated with different socio-demographics was also tested. Results indicated that in the selected context respondents younger than 25 years old were likely to be less rapidly satiated by personal activity in the morning compared to the people “over the 60s”. This could be due to parents’ or other household members’ sense of insecurity and unease associated with younger adults spending prolonged periods away from home for activities other than study or work in the afternoon and evening (Sharmeen and Houston, 2019). Further, people aged between 25 and 40 were less rapidly satiated to study in the morning. The strict schedule of educational institutes in Dhaka would persuade the respondents to spend longer time in the morning compared to the afternoon. People aged between 40 to 60 were more likely to spend longer time on work activities in the morning, compared to those “over the 60s”. Such findings can be explained by the fact that the retirement age in Bangladesh is the 60s. Therefore, it is plausible that respondents “over the 60s” did not need to start the office at the usual schedule (9:00-10:00). Also, respondents aged between 40 to 60 years were less rapidly satiated while accompanying compared to the people “over the 60s”. This is because respondents aged “over the 60s”, usually needs supervision for their health and well-being. Moreover, most of the educational institutions or health care hospitals located in Dhaka have opening time in the morning, which could induce the targeted respondents to spend longer time accompanying their dependants to the potential destination.

Along with the age of the respondents, gender effect on satiation was also tested in the final specification. Male respondents were found to have more rapid satiation for leisure activities compared to female when engaging in leisure activities in the morning. This finding aligns with the intuitive understanding that Bangladesh adheres to patriarchal norms where male individuals are typically the main earning person in the households. As a result, male individuals tend to maximise their utility by engaging more in productive activities to meet livelihood demands, leading to less time for pursuing discretionary activities. Major household responsibilities such as preparing meals, doing grocery shopping and accompanying children to school are often shouldered by females (Banks, 2013). Likewise, male respondents were likely to spend less time on shopping than female either starting the activity in the morning or afternoon. The prevailing social norms in the patriarchal societies of Bangladesh might have constrained female respondents from allocating more time to work or educational activities in comparison to their male counterparts (Lata et al., 2021). Furthermore, the shopping type (e.g., grocery, medicine, accessories, etc.) could have a significant influence on who engages in the shopping activity. As shopping types were not specified in the survey data, we were unable to conduct further investigations into the particular shopping activities undertaken by female respondents on weekdays, both in the morning and afternoon.

Like the other individual factors, respondents’ occupational status significantly influenced the satiation of performing an activity. The results from this study suggested that unemployed people were likely to be less rapidly satiated while accompanying others compared to employed personnel. On a working day, an employed person would have given the highest priority to work and spend longer time on work. On the other hand, unemployed respondents were likely to spend a longer time accompanying in the morning.

Impact of household sociodemographic on satiation

Besides the individual factors, the effect of household covariates on activity time use decision was also tested to capture the heterogeneity in satiation across decision makers. Respondents owning a personal vehicle (car/motorcycle) were more rapidly satiated to work in the morning, implying that they spent less time while started working before noon compared to the people from a household without a personal vehicle. In Bangladesh, vehicle ownership resembles higher social status, and this status insinuates a higher wage rate, and a higher position in the corporate office compared to the other. Holding a higher position means less accountability for their actions, hence, they are less likely to be affected by the consequences of a schedule delay (Zannat et al., 2022). Therefore, this group of people could prefer to spend less time in the morning on work activities. Also, owning a car or motorcycle provides the owner with a high degree of flexibility to earn from multiple sources (e.g., motorcycle/car owner can work as a ride-hailing service driver or can use their vehicle for renting at a certain time of the day), which also could affect their satiation of work activity, and the effect was significantly different in the morning. Because of this, members of a household with a personal vehicle can work less in the morning while still being able to meet their necessities for a living by working fewer hours in the morning or by working in the second half of the day.

Further, household income status affected the degree of satiation for out-of-home activity decisions. The satiation for travelling was found relatively lower for the high-income respondents (>60,000 BDT) compared to the other income groups. Higher-income people could afford to spend longer time travelling irrespective of the time of the day. This could be because of their flexible working hours (e.g., possibilities to have reduced work hours, having the option to do personal activities or be accompanied by a supporting hand) as opposed to other income groups.

Table 4-8 Bounded MDCEV results - Translation parameters.

	Activity		**Estimate**	**Rob.t.rat.(0)**
Baseline gamma	Work	Morning	10.589	48.314
		After noon	5.687	28.534
	Education	Morning	4.368	22.426
		After noon	4.655	9.511
	Shopping	Morning	2.622	6.435
		After noon	3.142	8.320
	Personal activity	Morning	5.110	23.994
		After noon	3.176	17.352
	Accompany	Morning	1.129	0.579
		After noon	1.684	2.969
	Leisure	Morning	6.988	7.465
		After noon	3.771	6.450
	Travel	Morning	0.229	36.500
		After noon	0.255	36.674
Age <25	Personal activity	Morning	0.249	1.316
Age 25-40	Education	Morning	0.636	2.329
Age 40-60	Work	Morning	0.139	1.872
	Accompany	Morning	0.341	1.703
Gender = male	Shopping	Morning	-0.234	-1.042
		After noon	-0.501	-2.654
	Leisure	Morning	-0.968	-2.638
Unemployed	Accompany	Morning	0.822	3.830
Vehicle ownership	Work	Morning	-0.268	-2.717
High income >60,000 BDT	Travel	Morning	0.193	3.831
		After noon	0.232	5.224

4.4.3. Significance of time varying utility and time of day specific budget constraint

Regardless of the demand modelling framework (econometric or rule-based), it is a common practice to quantify the activity preference by modelling different activity-related choice dimensions for demand generation of different market shares, and then scheduling the activities based on the preference of agents (Cirillo and Axhausen, 2010; Nurul Habib, 2018). A typical econometric activity model estimates the (dis)utility associated with each activity, attaches it to each activity and uses the results for scheduling (Bhat et al., 2004). Although a classical, econometric activity type and duration model highlighted users' utility and satiation for some activities, such a model ignored the fact that preferences could vary depending on activity start time. It is certain that the effects of a specific increase or decrease in each activity demand will be overestimated or underestimated if activity start time in relation to utility and satiation is ignored. The amount of travel required by commuters in the morning peak and afternoon off-peak, for instance, will not have the same effect on the transportation system's infrastructure, services, and environment. Furthermore, considering the same (dis)utility and satiation values estimated for each activity type for activity timing, irrespective of the time-of-day considerations, could be misleading. To accurately generate activity demand for the implementation of an effective demand management strategy, it is crucial to consider activity start time when performing activity-based demand modelling.

In this study, we provided empirical evidence using data from Dhaka, Bangladesh that activity start time jointly modelled with the activity type and duration for both mandatory and discretionary activities provided more intuitive understanding on individuals decision making. Instead of relying on the classical MDCEV framework, we adopted a bounded MDCEV framework to account for the fact that activities starting later time of the day will have a natural upper limit on their time allocation budget. Thus, the use of bounded MDCEV for joint modelling enabled to understand the intricate interplay among the choices of activity type, duration, and start time and estimate preference and satiation for out-of-home activities (both mandatory and discretionary activities) across different time-of-the-day. The adopted framework has two crucial benefits: (1) allowing utility difference based on activity start time, (2) accommodating time of day specific budget constraints while the overall budget of 24 hours was met. Instead of using the classical MDCEV (without start time specific budget constraint) model accounting for utility differences based on activity start time, the application of bounded MDCEV showed very little improvement in the prediction accuracy. As discussed by Saxena et al. (2021), the MDCEV model with bounds boils down to the traditional MDCEV model if no observation in the estimation data attains the upper bound. Since all observed afternoon time use from Dhaka data were smaller than the upper bound, the parameter estimates from the bounded MDCEV model were approximately same as the classical MDCEV model. However, the application of bounded MDCEV ensured that no individual was predicted with time allocations higher than the imposed upper bound, thus making the model logically consistent. By accounting for activity start time and time-of-the-day specific budget constraints, the classical MDCEV model was expanded to emulate the start-time sensitive nature of activity decision. This temporal consideration in activity modelling opens avenues for more informed decision-making related to activity timing.

4.5. Policy insights

In addition to the methodological implications of addressing challenges related to aligning the start times of activities with changing budget constraints in the activity type and duration choice, the findings of this study hold significant importance for the specific case we examined. To the best of the authors' knowledge, this is the first study that focuses on Dhaka, Bangladesh, a rapidly growing megacity that aims to develop a robust model to predict activity type and duration preference highlighting the variations in people's activity behaviours due to different budget constraints at different times of the day. Table 4-9 summarises the major findings and policy insights from the findings of this study.

Table 4-9 Summary of findings from the case study and policy insights.

Major findings	**Policy implications**
Highest utility and less rapid satiation for mandatory activities performed in the morning	This study reveals that in the context of Dhaka, people attained the highest utility from engaging in mandatory activities such as work and education in the morning. This insinuates a substantial need for morning travel to participate in these essential activities. However, these activities are often inflexible and cannot be easily postponed or rescheduled, especially if workplaces and educational institutions have limited flexibility. Additionally, less rapid satiation for mandatory activities performed in the morning highlighted that these activities need to be started in the morning due to their longer time requirement compared to the other activities. These findings align with the research by Zannat et al. (2024) which empirically demonstrates the significance of the correlation between outbound and duration of work trips. Therefore, special attention should be given to addressing the concentrated morning demand for mandatory activities, by implementing measures such as flexible working hours, and promoting remote work options (not yet very popular in Dhaka).
Higher utility for morning discretionary activities	In addition to mandatory activities, the study unveiled the higher utility among the individuals for specific discretionary activities i.e., accompanying in the morning. The increasing demand for accompanying in the morning carries significant policy implications, encompassing two key dimensions. Firstly, like the mandatory activities, accompanying someone introduces inherent challenges in postponement which has received less attention in the literature (Truong et al., 2017). These activities are contingent upon fixed schedules and institutional constraints, exemplified by commitments such as a child's school attendance or appointments for older individuals. An in-depth examination of these activities becomes imperative for mitigating peak-hour travel demand. Targeting institutions associated with such activities through policy measures is crucial. Secondly, effective demand management necessitates a nuanced understanding of the synchronised destination (e.g., school or day-care located approximately in the same location), schedule (e.g.,

	school start time 7:00 and office start time 9:00), and route characteristics (e.g., school and office located in the same route) of these activities. This insight becomes instrumental in devising policies that embrace a shared mobility approach, thereby contributing to the alleviation of peak-hour demand pressures. Similar to mandatory activities, less rapid satiation for personal and leisure activities also indicated a relatively longer time investment while performed in the morning, contributing to the need for an earlier start in the day. The correlation between the activity's start time and the duration of these discretionary activities reinforces the trend of morning preference for multiple alternatives. To alleviate peak-hour demand, it is essential to target these discretionary activities (Hoffman et al., 2013).
Utility and satiation variance across heterogeneous decision-makers (individual factors)	This study also highlighted the heterogeneity in activity preferences based on age, gender, and employment status in the selected context. It was found that respondents under 25 years old prioritised education, while those aged 25-40 were more focused on work. Moreover, young adults in Dhaka were less inclined to out-of-home discretionary activities on weekdays, indicating a trend towards a more sedentary lifestyle. This observation is also consistent with findings from Uddin (2019) in the same context. Gender also played a significant role in predicting the preference for out-of-home activities, with male respondents valuing mandatory activities the most, while female respondents derived the highest utility from personal activities and accompanying. This gender effect influenced by societal norms, which tend to place more earning responsibilities on males, can be observed in other cities from the Global South (Calastri et al., 2017). Furthermore, this study revealed the differences in satiation across male respondents' discretionary activities, which also supports the gender effect on activity and time preferences. Additionally, employment status affected the activity and time use decision. Unemployed individuals, not constrained by institutional schedules, showed a higher likelihood for morning periods for discretionary activities. To address these varying preferences and demands, it is essential to consider differences in activity patterns related to age, gender, and employment status. Each group achieved their highest utility from different activities, depending on their level of flexibility. This underscores the importance of tailoring strategies and policies to meet the diverse needs of the population.
Utility and satiation variance across heterogeneous decision-makers	This study highlighted the influence of factors like the presence of children in households, household vehicle ownership, income, and household location on activity preferences throughout the day. Consequently, in comprehending the intricate patterns of activity-travel behaviour, it is

(household factors)	essential to account for the diversity stemming not only from individual factors but also from household demographics (Pinjari and Bhat, 2011). In the context of Dhaka, it is timely to take measures acknowledging the household diversity such as improving childcare services to support working parents and ensuring accessible and affordable leisure activities, particularly for middle- and low-income groups.

4.6. Conclusion

The aim of this paper is to investigate the intricate interplay among the choices of activity type, duration, and start time that influence the activity decision pertaining to both mandatory and discretionary activities. While accommodating activity start time in the activity model, this study goes further the Pinjari and Bhat (2010) by not only accounting for utility differences depending on activity start time, but also by explicitly accommodating time of day specific time budget constraints. Through the application of the bounded MDCEV model, the authors attempted to jointly model activity type, start time and duration selecting Dhaka, Bangladesh as a case study. Results highlighted the heterogeneity in activity participation in the context of Dhaka. Different socio-demographic factors emerged as significant contributors to the intricate relationships among activity types, their durations, and the choices of when to initiate these activities. Moreover, it is evident from the results of this study that decision makers' satiation and utility associated with out-of-home activity participation varied with the activity start time depending on decision makers' individual and household characteristics (age, gender, employment status, vehicle ownership, income, home location, etc.). Therefore, the estimation of a generic parameter independent of activity starting time would be misleading for forecasting. From a policy perspective, the use of activity timing information jointly with activity type choice and duration provides rich situational information about activity and travel patterns.

Though this study attempted to shed light on the importance of the consideration of activity start time in the activity type and duration choice model, however, the study can be further extended. To avoid identification issues[24], the authors divided the timing decision into two coarse segments (morning and afternoon). Depending on the purpose of the study and data quality, activity starting time can be divided into the finer scale (e.g., morning peak, morning off-peak, evening peak, evening off-peak, etc.) along with precise (and appropriate) upper (and lower) bounds for activity types. Further, the authors did not test the importance of timing preference with a detailed classification of at-home activities. Additionally, for multiple occurrences of activity, in future research, it is important to consider the time interval between two consecutive activities in activity time preference and time use decision. Since this study was done using 24-hour one-day survey information, we could not capture the unobserved correlation among activities performed in the morning and afternoon. Therefore, the current model can be further improved by using panel data to capture the correlation between alternatives, and inter-intra heterogeneity associated with utility and satiation of activity start time and time use decision.

24 An increase in the number of possible time periods could contribute to an increase in the number of possible time periods and activity pairings, which could increase the number of constants, a large number of unchosen alternatives, and correlation among the alternatives. Eventually, this would then exacerbate the problems with computational cost and parameter identification.

Also, instead of relying on conventional travel diary survey data, this study can be further extended by using passive data. For example, using smart card information or high-resolution GPS data, it is feasible to anticipate probable activity starting time and activity duration of respondents for a longer duration (Zannat and Choudhury, 2019). However, even in its current form, the research findings can be practically useful to test the impact of different activity starting times, opening times, latest start time and earliest end time of different activity places within an agent-based simulation platform. Eventually, the outcome from the final specification can be used for activity timing in an agent-based simulation models, which would be useful to understand the impact of individuals' activity decisions at the macro scale (at the neighbourhood or city level).

4.7. Reference

Andaleeb, S.S. and Vanneman Wolford, G. 2004. Participation in the workplace: gender perspectives from Bangladesh. *Women in Management Review.* **19**(1), pp.52-64.

Axhausen, K.W., Horni, A. and Nagel, K. 2016. *The multi-agent transport simulation MATSim.* Ubiquity Press.

Balać, M., Janzen, M. and Axhausen, K.W. 2018. Alternative approach to scoring in matsim and how it affects activity rescheduling. In: *2018 TRB Annual Meeting Online*: Transportation Research Board, pp.18-02016.

Banks, N. 2013. Female employment in Dhaka, Bangladesh: participation, perceptions and pressures. *Environment and urbanization.* **25**(1), pp.95-109.

BBS. 2011. *Bangladesh Population and Housing Census*

Bhat, C.R. 1996. A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity. *Transportation Research Part B: Methodological.* **30**(3), pp.189-207.

Bhat, C.R. 2005. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological.* **39**(8), pp.679-707.

Bhat, C.R. and Koppelman, F.S. 1993. A conceptual framework of individual activity program generation. *Transportation Research Part A: Policy and Practice.* **27**(6), pp.433-446.

Bhat, C.R. and Koppelman, F.S. 1999. A retrospective and prospective survey of time-use research. *Transportation.* **26**(2), pp.119-139.

Bhat, C.R., Guo, J.Y., Srinivasan, S. and Sivakumar, A. 2004. Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record.* **1894**(1), pp.57-66.

Bowman, J.L. and Ben-Akiva, M.E. 2001. Activity-based disaggregate travel demand model system with activity schedules. *Transportation research part a: policy and practice.* **35**(1), pp.1-28.

Calastri, C., Hess, S., Daly, A. and Carrasco, J.A. 2017. Does the social context help with understanding and predicting the choice of activity type and duration? An application of the Multiple Discrete-Continuous Nested Extreme Value model to activity diary data. *Transportation Research Part A: Policy and Practice.* **104**, pp.1-20.

Cirillo, C. and Axhausen, K.W. 2010. Dynamic model of activity-type choice and scheduling. *Transportation.* **37**(1), pp.15-38.

Eluru, N., Pinjari, A., Pendyala, R. and Bhat, C. 2010. An econometric multi-dimensional choice model of activity-travel behavior. *Transportation Letters.* **2**(4), pp.217-230.

Enam, A. and Choudhury, C.F. 2011. Methodological issues in developing mode choice models for Dhaka, Bangladesh. *Transportation research record.* **2239**(1), pp.84-92.

Flavia, A. and Choudhury, C. 2019. Temporal transferability of vehicle ownership models in the developing world: case study of Dhaka, Bangladesh. *Transportation Research Record.* **2673**(3), pp.722-732.

Golshani, N., Shabanpour, R., Auld, J. and Mohammadian, A. 2018. Activity start time and duration: incorporating regret theory into joint discrete–continuous models. *Transportmetrica A: transport science.* **14**(9), pp.809-827.

Habib, K.M., Carrasco, J.A. and Miller, E.J. 2008. Social context of activity scheduling: Discrete–continuous model of relationship between "with whom" and episode start time and duration. *Transportation Research Record.* **2076**(1), pp.81-87.

Habib, K.M.N. 2011. A random utility maximization (RUM) based dynamic activity scheduling model: Application in weekend activity scheduling. *Transportation.* **38**(1), pp.123-151.

Habib, K.M.N. 2012. Modeling commuting mode choice jointly with work start time and work duration. *Transportation research part A: Policy and practice.* **46**(1), pp.33-47.

Habib, K.M.N. and Carrasco, J.-A. 2011. Investigating the role of social networks in start time and duration of activities: Trivariate simultaneous econometric model. *Transportation research record.* **2230**(1), pp.1-8.

Habib, M.A. and Daisy, N.S. 2013. Modeling frequency and duration of out-of-home participation in physical activity by school-age children. *Transportation research record.* **2357**(1), pp.116-124.

Hoffman, K., Berardino, F. and Hunter, G. 2013. Congestion pricing applications to manage high temporal demand for public services and their relevance to air space management. *Transport Policy*. **28**, pp.28-41.

Khaddar, S., Saxena, S., Fatmi, M.R. and Pinjari, A.R. 2022. An Episode-Level Joint Model for Activity Engagement, Duration, and Destination Location Choice Decisions. *Transportation Letters.* pp.1-11.

Khan, N.A., Enam, A., Habib, M.A. and Konduri, K.C. 2022. Exploring joint activity-tour participation, time allocation, and mode choice decisions. *Transportation Letters.* pp.1-11.

Lata, L.N., Walters, P. and Roitman, S. 2021. The politics of gendered space: Social norms and purdah affecting female informal work in Dhaka, Bangladesh. *Gender, Work & Organization.* **28**(1), pp.318-336.

Liu, C., Susilo, Y.O. and Karlström, A. 2017. Jointly modelling individual's daily activity-travel time use and mode share by a nested multivariate Tobit model system. *Transportmetrica A: Transport Science.* **13**(6), pp.491-518.

Nurul Habib, K. 2018. A comprehensive utility-based system of activity-travel scheduling options modelling (CUSTOM) for worker's daily activity scheduling processes. *Transportmetrica A: Transport Science.* **14**(4), pp.292-315.

Palma, D., Enam, A., Hess, S., Calastri, C. and Crastes dit Sourd, R. 2021. Modelling multiple occurrences of activities during a day: an extension of the MDCEV model. *Transportmetrica B: Transport Dynamics.* **9**(1), pp.456-478.

Pawlak, J., Polak, J.W. and Sivakumar, A. 2017. A framework for joint modelling of activity choice, duration, and productivity while travelling. *Transportation Research Part B: Methodological.* **106**, pp.153-172.

Pinjari, A.R. and Bhat, C. 2010a. An efficient forecasting procedure for Kuhn-Tucker consumer demand model systems. *Technical paper. Department of Civil & Environmental Engineering, University of South Florida.*

Pinjari, A.R. and Bhat, C. 2010b. A multiple discrete–continuous nested extreme value (MDCNEV) model: Formulation and application to non-worker activity time-use and timing behavior on weekdays. *Transportation Research Part B: Methodological.* **44**(4), pp.562-583.

Pinjari, A.R. and Bhat, C.R. 2011. Activity-based travel demand analysis. *A handbook of transport Economics.* **10**, pp.213-248.

Rahman, H.Z. 2016. *Bangladesh 2016: Politics, Governance, and Middle Income Aspirations Realities and Challenges* Dhaka, Bangladesh: : Power and Participation Research Centre (PPRC).

Saxena, S., Pinjari, A.R., Roy, A. and Paleti, R. 2021. Multiple discrete-continuous choice models with bounds on consumptions. *Transportation Research Part A: Policy and Practice.* **149**, pp.237-265.

Sharmeen, N. and Houston, D. 2019. Spatial characteristics and activity space pattern analysis of Dhaka city, Bangladesh. *Urban Science.* **3**(1), p36.

Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Bjelland, J., Engø-Monsen, K., De Montjoye, Y.-A. and Iqbal, A.M. 2017. Mapping poverty using mobile phone and satellite data. *Journal of The Royal Society Interface.* **14**(127), p20160690.

Timmermans, H. 2005. *Activity-based approaches: models, data and applications.*

Truong, L.T., De Gruyter, C., Currie, G. and Delbosc, A. 2017. Estimating the trip generation impacts of autonomous vehicles on car travel in Victoria, Australia. *Transportation.* **44**, pp.1279-1292.

Uddin, R. 2019. Physical activity, sedentary behaviour and psychological wellbeing among young adults in Bangladesh: A one-year prospective study.

Varghese, V. and Jana, A. 2019. Multitasking during travel in Mumbai, India: Effect of satiation in heterogeneous urban settings. *Journal of Urban Planning and Development.* **145**(2), p04019002.

Wadud, Z. 2020. The effects of e-ridehailing on motorcycle ownership in an emerging-country megacity. *Transportation research part A: policy and practice.* **137**, pp.301-312.

WPR. 2021. *World Population Review - Dhaka population 2021*. [Online]. [Accessed 5 September]. Available from: https://worldpopulationreview.com/

Wu, G. and Hong, J. 2022. An analysis of the role of residential location on the relationships between time spent online and non-mandatory activity-travel time use over time. *Journal of Transport Geography.* **102**, p103378.

Zannat, K.E. and Choudhury, C.F. 2019. Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science.* **99**(4), pp.601-619.

Zannat, K.E., Choudhury, C. and Hess, S. 2024. Modelling Time-of-Travel Preferences Capturing Correlations between Departure Times and Activity Durations. *Transportation Research Part A: Policy and Practice.*

Zannat, K.E., Choudhury, C.F. and Hess, S. 2022. Modeling departure time choice of car commuters in Dhaka, Bangladesh. *Transportation research record.* **2676**(2), pp.247-262.

Zhang, Y. and Yao, E. 2022. Exploring elderly people's daily time-use patterns in the living environment of Beijing, China. *Cities.* **129**, p103838.

Appendix B

Supplementary Table 4-1 Occupation wise activity participation summary.

	Student	Public employee	Private employee	Business	Agriculture	Housewife	Unemployed	Retired	Other
Work	0.12	9.29	48.52	39.26	0.26	0.04	0.03	0.03	0.06
Studying	99.07	0.08	0.37	0.17	0.01	0.25	0.03	0.00	0.03
Shopping	7.43	3.69	12.26	13.78	1.89	53.65	2.37	3.98	6.34
Personal activity	10.59	3.91	13.50	18.49	1.08	38.27	3.53	8.97	12.50
Leisure and social	18.92	5.20	13.31	13.72	0.42	37.42	2.70	7.48	10.19
Accompanying	1.57	0.75	2.48	5.96	0.08	86.51	1.16	1.16	2.32
Travel	28.48	5.26	26.97	23.25	0.41	11.75	0.70	1.60	2.31

Chapter 5 The role of satiation in activity participation: a comparative analysis of fractional logit (FMNL) model and multiple discrete-continuous extreme value (MDCEV) model

Abstract

In activity modelling, the concept of satiation, which reflects the diminishing marginal benefits of engaging in a particular activity over time, has received significant attention. Despite the dependency of satiation on context, wage rate, family structure, and exposure pattern, the logarithmic function has been uniformly applied for all activities as the utility of activity duration in activity modelling. However, according to McAlister (1982), satiation is not a permanent state, yet its impacts on activity modelling and predicting activity duration have been rarely investigated. This is particularly important when considering activity modelling in the context of the Global South, where individuals often contend with limited access to resources, tend to prioritise activities essential for meeting their basic needs, survival, or livelihood. Therefore, the aim of this study is to investigate the impact of incorporating satiation into activity choice modelling on the accuracy of predicting activity durations, examine how accounting for correlation affects the findings regarding satiation in predicting activity duration, and eventually, assess potential gain in model improvement after incorporating random heterogeneity in time use models. We estimated activity models (with and without satiation) using a detailed multi-day travel diary survey data from 170 residents of the Greater Concepción Area, Chile. Results revealed that in this specific sample, where a majority of the respondents were full-time workers or students, FMNL outperformed the simple MDCEV model in recovering the average time use for different activities. Moreover, an improvement in the MDCEV model fit was evident after incorporating heterogeneity in baseline preferences and satiation while estimating activity model by utilising latent class MDCEV framework. However, for the selected sample profile heterogeneity in baseline preferences outperformed the heterogeneity in satiation in improving the model fit. Capturing correlations among the alternatives using the multiple discrete continuous nested extreme value model (MDCNEV) also resulted in model improvements and better predicted the average time use of the decision makers if participated. Findings from this study suggested that to improve the model fit and prediction accuracy of the activity model, it is necessary to investigate the varied specification effect of satiation on activity duration. The results can guide transport planners to make a more informed choice about the appropriate framework for modelling activity choice and time allocation.

Keywords: Satiation, Latent class model, activity type and duration choice model, fractional logit model (FMNL), multiple discrete-continuous extreme value model (MDCEV), multiple discrete continuous nested extreme value model (MDCNEV)

5.1. Introduction

Activity-based models are gaining increasing popularity among transport researchers and practitioners for characterising transport demand (Timmermans, 2005; Rasouli and Timmermans, 2012; You et al., 2016; Eom et al., 2020; Hamad and Obaid, 2022). Real-world activity decisions are driven by the simultaneous choice of multiple alternatives synergised with different activity dimensions such as type, duration, and timing. In activity modelling, the concept of satiation, which reflects the diminishing marginal benefits of engaging in a particular activity over time, has garnered significant attention (Bhat, 2008; Pinjari et al., 2009; Calastri et al., 2017; Varghese and Jana, 2019). This is apparent in the activity/agent-based modelling that the logarithmic function was uniformly applied for all activities as the utility of activity duration (Axhausen, Kay W. et al., 2016). While the marginal utility of activity participation may decline over time for some activities, it may not be applicable for a variety of activities (such as long waited leisure and social activities) and can be influenced by socio-economic conditions. This is particularly important when considering activity modelling in the context of the Global South, especially in the least developed countries, where individuals often contend with limited access to resources. In such cases, people tend to prioritise activities essential for meeting their basic needs, survival, or livelihood, resulting in fewer opportunities for engagement in activities that could lead to satiation. Also, it is evident from empirical investigation that people tend to work until they attain a particular target of income and satiation is associated with the wage rate (Niyogi, 2015). Furthermore, distinct family structures, including household size, extended family dynamics, and the presence of sole breadwinners, contribute to ongoing struggles within both formal and informal labour sectors to fulfil daily necessities, even with extended work hours. Prolonged periods of work may yield additional earnings for these families, affording them the chance to provide better meals or indulge in long-awaited leisure activities.

The satiation effect also depends on the variations in exposure patterns. When individuals engage repeatedly in a preferred activity, a decline in the enjoyment or satisfaction derived from it due to satiation may occur. However, the diminishing satisfaction is not a permanent state (McAlister, 1982). Over time, and with a break or change in the pattern of consumption, individuals may recover or reset, making the stimulus enjoyable again (Galak et al., 2011). From conventional travel diary survey data, which typically collects aggregated activity information such as work, education, and social activities, it is challenging to understand how individuals organise their daily activities—both in-home and out-of-home—to manage satiation through involvement in multiple alternatives within time budget constraints (Axhausen, Kay Werner, 2007). Furthermore, the motivation of decision-makers to sustain prolonged engagement without experiencing satiation is influenced by their past experiences and tolerance levels (McSweeney and Swindell, 1999). For instance, an individual's initial experience of an 8-hour workday in a new office setting may induce fatigue, but with continued exposure to the workplace culture, their response to an 8-hour workday may adapt. This adaptation illustrates the role of past experiences and individual tolerance levels in shaping responses to extended durations of activity. In addition, satiation could be person specific (Kapur and Bhat, 2007). For example: an overtime shift might not have the same satiation effect on a workaholic as it would on someone who is less industrious. Therefore, if satiation is inappropriately incorporated into the model, it may lead to either an overestimation or underestimation of the time use preference. This discrepancy may arise if the influence of satiation is not thoroughly examined within the context of the chosen sample profile, which exhibits a heterogeneous time use pattern.

In the field of psychology and marketing research, scholars delve into different methodological approach to investigate the impact of satiation on consumption choices, considering a spectrum of engagement-related factors such as variety, the repetition of participation, and exposure (McAlister, 1982; McSweeney and Swindell, 1999; Redden, 2008; Park and Jang, 2014; Line et al., 2016). The exploration of integrating satiation into activity modelling, whether explicitly or implicitly, and its potential impact on the accuracy and behavioural insights in predicting activity demand is significantly lacking in existing literature (Chen and Liao, 2019). To address this gap, this study took the initial step of analysing the implications of considering satiation when modelling activity type and duration choice with a focus on a case study from the Global South — Concepción, Chile. This study sought to answer the following research questions: 1) How does incorporating satiation into activity choice modelling affect the accuracy of predicting activity durations? 2) How does accounting for correlation affect the findings regarding satiation in predicting activity duration, 3) Can additional insights be gained by incorporating random heterogeneity in time use models? Therefore, the aim of this study is to investigate the impact of incorporating satiation into activity choice modelling on the accuracy of predicting activity durations, to examine how accounting for correlation affects the findings regarding satiation in predicting activity duration, and to assess potential gain in model improvement after incorporating random heterogeneity in time use models.

In contemporary literature, different modelling frameworks have been developed to accommodate the diverse dimensions of activity choices (such as activity type, duration, and start time), encompassing both discrete and continuous aspects. These approaches include Lee's approach (Lee, 1982), which transforms the error terms of discrete and continuous dimensions to adhere to a bivariate normal distribution (Habib, 2012). Despite the popularity of Lee’s approach in discrete-continuous joint modelling, its strict imposition of bivariate normal coupling can lead to restrictions and result in inflexible structures (Spissu et al., 2009). Additionally, copula-based methodologies are employed, allowing for the relaxation of specific distributional assumptions and enabling the capture of multivariate dependency structures between error terms (Golshani et al., 2018; Shabanpour et al., 2018). While the copula-based approach is powerful for modelling the dependence structure between random variables, as the number of variables increases, the computational burden of estimating the copula parameters and performing inference grows significantly. Other methods include the multivariate probit model and multinomial poisson approach renowned for modelling multiple discreteness, however, lack a rigorous underlying utility maximisation framework (Terza and Wilson, 1990; Bhat and Srinivasan, 2005). The multiple discrete continuous extreme value model (MDCEV), introduced by (Bhat, 2008) accommodates both discrete and continuous dimensions of activity choices and has played a significant role in advancing the understanding of activity choice modelling over the past two decades (Bhat, 2005; Bhat et al., 2006; Kapur and Bhat, 2007; Rajagopalan et al., 2009; Pinjari and Bhat, 2010; Bernardo et al., 2015; Calastri et al., 2017; Palma et al., 2021; Saxena et al., 2021). The MDCEV model has predominated the current activity research due to the flexibility to accommodate ‘satiation’. Another potential model is the fractional logit model (FMNL) (Papke and Wooldridge, 1996), which can handle the simultaneous choice of multiple alternatives, however, very rarely applied in the context of activity choice modelling (Ye and Pendyala, 2005). The fractional logit model relies on a quasi-maximum likelihood estimator which does not necessitate a full normal distributional assumption for consistency (Cardoso et al., 2010). In our investigation, we have chosen to focus on two potential candidate models: the MDCEV and the Fractional logit model. Both models have the capability to predict activity duration, accommodate multiple alternatives, and capture non-linear effects. However, it is important to note that while we have selected these two candidate models for examination, other models could also be explored.

Three different approaches were adopted: (1) model fit from FMNL and MDCEV models were compared respectively to assess which utility specification yielded better prediction accuracy of activity duration, while the former one predicted time use where the prediction of mean time use fall within the logit curve (non-linear and non-additive prediction for proportion) (Dubin, 2007; Becker, 2014; Murteira and Ramalho, 2016; Kochaniak, 2020), and latter one explicitly accommodated satiation and enabled to estimate satiation parameters (non-linear and additive form of utility) (Bhat, 2005), (2) the findings from satiation in predicting activity duration was further assessed while accounting for correlation among the alternatives, and (3) the additional gain in the model fit was compared after capturing the unobserved heterogeneity using the latent class FMNL and latent class MDCEV model. The main contribution of this study is twofold — (1) the role of satiation consideration was analysed in the activity time choice modelling (with/without correlation among the alternatives), and (2) insights gained by incorporating random heterogeneity in time use models was investigated.

The rest of the paper is organised as follows. In section two, a short overview of the survey and data collection is given along with the sample statistics. In the next section, a description of the modelling framework is given which is followed by the model estimation. Eventually, the results from the model estimation are then presented with a critical discussion of this study.

5.2. Data

5.2.1. Survey and data collection

In this study, we used detailed travel diary survey data accumulated between 2015 and 2016 from the residents of the Greater Concepción Area, Chile. Concepción is the regional capital of the Greater Concepción metropolitan area, the third largest conurbation in Chile. It is located approximately 500 km south of the capital, Santiago. The survey involved respondents from six different neighbourhoods located in the northeastern part of the city. Each neighbourhood was selected by its location and income status. Geographically, all the selected neighbourhoods have homogeneous land use patterns and relatively similar connectivity with the central business district (CBD). Income-wise, among the selected neighbourhoods there are two low-income, two middle-income, and two high-income neighbourhoods (Figure 5-1). The adopted sampling followed a quota sampling procedure based on the available census information considering age, gender, and occupation.

The data were collected through a face-to-face interview. In the first section of the survey, participants were requested to complete a questionnaire related to their personal and household socio-demographics, including questions about age, gender, occupation, educational qualification, marital status, household income, car ownership, housing type, driving license, the presence of children and other members at home, and household size. Then, the respondents were asked to complete a seven-day activity diary with detailed information on their activity participation behaviour. The survey recorded the description of each activity, trip-related information to participate in the corresponding activity (departure time, travel time, mode), activity starting and end time, location information where the activity was carried out, and with whom the activity was performed. A very short description of the survey sample is given in this article for brevity. A detail description of the survey method and data summary is available in Victoriano et al. (2020).

In addition to the activity diary, using the technique "name generator", proposed by Carrasco et al. (2008), respondents were asked about their social networks. Respondents reported about the people they considered emotionally very close and somewhat close to them. Such an approach is useful for understanding communication and social activity-travel patterns. More details about the method can be found in Carrasco et al. (2008).

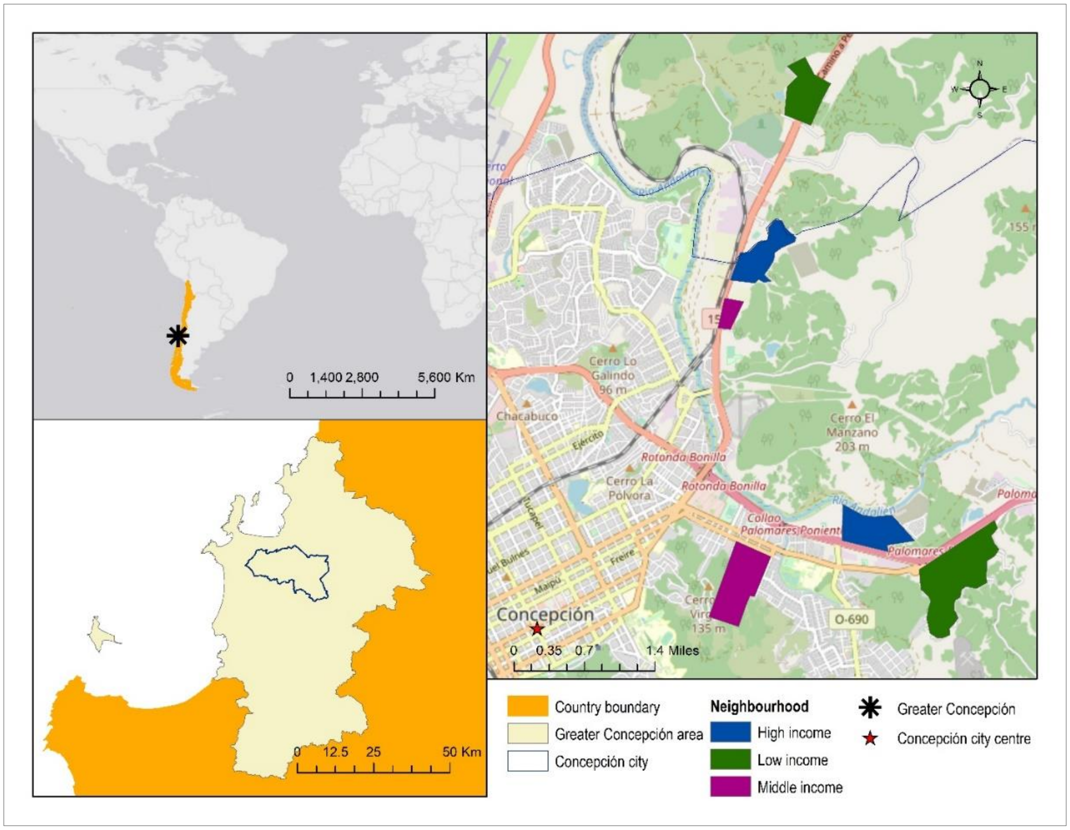


Figure 5-1: Location map of study area.

5.2.2. Sample characteristics

A total of 170 respondents were involved in the multiday interview, resulting in a database of 1187 days and 16376 activities and trips. Table 5-1 summarises the socio-demographic characteristics of respondents. Within the selected sample, the proportion of female was higher than the male (1:1.6), and the highest proportion of respondents was aged between 40-60 years old i.e., the working adult group. The sample consisted of an almost equal proportion of respondents from different income groups and all the respondents had at least basic schooling. The average household size in the sample was about 4. The majority of the respondents were married or living with partners (68%). Approximately, 68% of the respondents had at least one car and 44% with driving license. Households had children 15 years old or less in 51% of the cases.

Table 5-1 Socio-demographic characteristics of the sample.

	N= 170
Individual Characteristics	
Gender	
Male	38.24%
Female	61.76%
Age	
< 26 years	15.29%
26 – 40 years	35.88%
40 – 60 years	41.76%
>= 60 years	7.07%
Average monthly household income (CL$' 000)*	
<= 400	30.58%
400– 1000	36.46%
> 1000	32.96%
Education level	
Base	8.82%
Primary	34.12%
Secondary	20.00%
Under/Grad.	37.06%
Car ownership	
Yes	68.83%
No	31.17%
Driving license	
Yes	55.88%
No	44.12%
Occupation	
Working or Studying	72.35%
Other	27.65%
Household Characteristics	
Household size	Mean 4.15 SD (1.54)
Having young children	
Yes	51.18%
No	48.82%
Married / Have partner	
Yes	68.82%
No	31.18%
Occupation of partner	
Working	51.18%
Not working	48.82%

* 1000 CL$≈1.2 USD

As we modelled the choice and duration of activities, the time use diary was the core part of the dataset for our analysis. In the case of reported activities exceeding 24 hours a day, we removed/shortened/reattributed the last activities reported. The choice set of the activity agenda is defined by 14 individual activity types including both in-home and out-of-home activities.

Activity type 1: *Basic needs* included sleep/nap, getting dressed/ready, and personal accessories. Even if the person was out-of-home, he/she had to do such activities wherever they were staying over (staying with parents, friends, out of Concepción).

Activity type 2: *Household obligatory* activities included cooking/cleaning, breakfast/lunch/dinner, taking care of children/pets/older people, and family time. Eating lunch and dinner alone or with family members was included as part of this type of activity.

Activity type 3: *In-home work* activities included office work/meetings doing at home, and preparation for a job.

Activity type 4: *In-home education* included doing assignments/lessons for school/university, and preparation for exams.

Activity type 5: *In-home leisure* activities included reading novels, watching TV, browsing the internet/social media, hobbies, and exercises performed at home.

Activity type 6: *In-home social* activities included gatherings or dinners with friends/neighbours/relatives at home.

Activity type 7: *Work* activities included work activities or work related out-of-home activities (meeting with colleagues, attending conferences), participating in job interviews, etc. If the person was currently studying full-time (though he has a part-time job), then attending a conference was part of out-of-home education activity for that person.

Activity type 8: *Education* activities included performing education activities at school, library, friend's house, tutor's house, or any other out-of-home places.

Activity type 9: *Personal* activities included medical/hospital visits for health check-ups, and other institutional/organizational activities such as banking, paying bills, servicing automobiles, etc.

Activity type 10: *Shopping* activities for groceries, food, furniture, clothing, etc.

Activity type 11: *Leisure* activities included out-of-home leisure activities such as exercise, sports, cinema, going for walks, and other hobbies performed outside of home.

Activity type 12: *Social* activities included visiting family/friends, gathering with friends/colleagues at bars/clubs, planned social events (birthday party, marriage ceremony), etc.

Activity type 13: *Drop-off/Pick-up* activities such as dropping/picking up children, partners, etc.

Activity type 14: *Travel* includes travel time to reach the activity location.

Table 5-2 Heterogeneity in activity type and duration choice (weekdays and weekend) (observation summary).

Activity type	Weekdays		Weekend		Aggregated	
	% of observation engage in the activity	Average time spent (±SD) by those who choose it (h)	% of observation engage in the activity	Average time spent (±SD) by those who choose it (h)	% of observation engage in the activity	Average time allocation
Basic needs	100.00	11.27 (±2.78)	100.0	12.69 (±3.13)	100.00	11.68
Household obligation	55.31	3.2 (±2.47)	54.3	3.26 (±2.23)	55.01	1.77
In-home work	13.68	2.98 (±2.44)	8.0	3.69 (±2.85)	12.05	0.37
In-home education	4.25	3.61 (±2.18)	3.8	3.25 (±2.09)	4.13	0.15
In-home leisure	57.78	2.96 (±2.18)	56.9	3.17 (±2.43)	57.54	1.74
In-home social	14.50	1.85 (±1.60)	16.8	2.15 (±1.52)	15.16	0.29
Work	54.72	6.78 (±2.64)	20.4	5.35 (±3.16)	44.90	2.96
School	8.84	4.49 (±2.98)	2.4	3.6 (±3.33)	6.99	0.31
Personal issues	26.18	1.85 (±1.84)	11.8	1.75 (±2.29)	22.07	0.40
Shop	39.39	1.01 (±1.16)	44.2	1.28 (±1.11)	40.78	0.44
Leisure	11.20	2.5 (±1.91)	23.3	3.07 (±2.26)	14.66	0.41
Social	55.90	2.38 (±2.26)	59.3	4.36 (±2.88)	56.87	1.69
Accompany/Drop-off	23.47	0.92 (±0.77)	9.7	1.21 (±1.16)	19.55	0.19
Travel	93.04	1.73 (±1.31)	88.8	1.78 (±1.20)	91.83	1.60

5.3. Methodology

5.3.1. Modelling framework

To address our research questions, we selected two candidate models: the multiple discrete continuous extreme value model and the fractional Logit model. Both models offer flexibility in accommodating the simultaneous choice of multiple alternatives. In the following subsection, the principles, and modelling framework of the selected models are explained.

5.3.1.1. Fractional multinomial logit model (FMNL)

Time spent on travel and other activities can be thought of as fractions of the overall daily time resource that is accessible to an individual. A quasi-maximum likelihood method known as the fractional logit model is convenient for modelling such fractional dependent variables. Here, in the situation of multiple alternatives with non-zero outcome $0 \leq x_{j,n,t} \leq 1$ and $\sum_{j=1}^{J} x_{j,n,t} = 1$ where *n* (*n=1, 2, 3 …., N* and *N* is the sample size) is an index that represents the individual and *j (j = 1, 2, 3 …., J)* is an index that represents the alternatives in the choice set. By taking into account the product across all the non-zero alternatives, the probability of individual *n* choosing *M* of the *J* alternatives in task *t*, conditional on utility parameters can be formulated as follows:

$$P_{j,n,t}(\beta) = \frac{\exp\,(\sum_{m=1}^{M} x_{i,n,t}\, V_{i,n,t})}{\sum_{j=1}^{J} \exp\,(V_{j,n,t})} \tag{1}$$

Here, V is the observed utility and β is the parameters (related to the activity type and characteristics of the decision maker) to be estimated. Therefore, the likelihood of the observed sequence of choices for person n is then given by:

$$L_n(\beta) = \prod_{t=1}^{T_n} \frac{\exp\,(\sum_{m=1}^{M} x_{i,n,t}\, V_{i,n,t})}{\sum_{j=1}^{J} \exp\,(V_{j,n,t})} \tag{2}$$

5.3.1.2. Fractional nested logit model (FNL)

Alternative activities are likely to have unobserved correlations in the error terms. To capture the correlation among the alternatives in the fractional model, we used nested logit (NL) formulation. In the NL model correlations are allowed among alternatives (Koppelman and Wen, 1998; Hensher et al., 2008). Probability of 14 alternatives were estimated for different nesting structures. The probability of individual n choosing alternative i in choice situation t where i belongs to nest S_m :

$$P_{i,n,t} = P_{S_m} P_{i|S_m} \tag{3}$$

Here, the nest probability P_{S_m}:

$$P_{S_m} = \frac{e^{\lambda_m I_m}}{\sum_{m=1}^{M} e^{\lambda_I I_I}} \tag{4}$$

The probability of choosing alternative i belongs to nest S_m:

$$P_n(i|S_m) = \frac{\exp\,(\frac{V_{i,n,t}}{\lambda_m})}{\sum_{j \epsilon S_m} \exp\,(\frac{V_{j,n,t}}{\lambda_m})} \tag{5}$$

$$I_m = ln \sum_{j \epsilon S_m} \exp\,(\frac{V_{j,n,t}}{\lambda_m}) \tag{6}$$

The Likelihood function of the observed sequence of choices for person n is given by:

$$L_n(\beta) = \prod_{t=1}^{T_n} \prod_{j=1}^{J} (P_{j,n,t})^{x_{j,n,t}} \tag{7}$$

5.3.1.3. Multiple discrete–continuous extreme value model (MDCEV)

Another utility maximization theory-based econometric model, known as the multiple discrete–continuous extreme value (MDCEV) framework, and its extended versions have been used in recent studies to account for the multiple discreteness as well diminishing marginal returns (i.e., *satiation*).

The utility equation introduced by Bhat (2008) is given as follows:

$$U(x) = \frac{1}{\alpha_1}\psi_1 {x_1}^{\alpha_1} + \sum_{j=2}^{J} \frac{\gamma_j}{\alpha_j}\psi_k \left(\left(\frac{x_j}{\gamma_j}+1\right)^{\alpha_j} - 1\right) \quad (8)$$

U(x) is quasi-concave and continuously differentiable with respect to the consumption quantity vector *x*. ψ_j is the baseline utility of good *j* i.e., the marginal utility of the good at the point of zero consumption. ψ_j is the function of observed characteristics of decision maker and good *j*. ψ_j can be further parameterized as:

$$\psi_j = e^{\delta_j + \beta_j * z_j + \in_j} \quad (9)$$

δ_j is constant for alternative *j* for the baseline utility, z_j is the vector of the attribute associated with the alternative (activity type) and characteristics of the decision maker. $\in_j$ is the unobserved characteristics that affect the baseline utility for activity *j*. α_j and γ_j are related to activity *j*. The role of α_j is to reduce the marginal utility with increasing consumption of good *j*. γ_j is the translation parameter that allows for corner solution and is associated with the level of satiation. In our MDCEV model, we used generic α with alternative specific γ_j as it leads to a more efficient forecasting algorithm (Pinjari and Bhat, 2010). Hence, γ_j can be further parameterized as:

$$\gamma_j = \varphi_j + \lambda_j * Z_j \quad (10)$$

φ_j is constant for alternative *j* and λ_j is estimated parameters capturing the impact of Z_j. The probability that an individual chooses a specific vector of consumption amount:

$$P({x_1}^*, {x_2}^*, \dots\dots, {x_M}^*, 0, \dots, 0) = \frac{1}{P_1}\frac{1}{\sigma^{M-1}}\left(\prod_{m=1}^{M} f_m\right)\left(\sum_{m=1}^{M}\frac{P_m}{f_m}\right)\left(\frac{\prod_{m=1}^{M} e^{V_i/\sigma}}{\left(\sum_{j=1}^{J} e^{V_j/\sigma}\right)^M}\right) \quad (11)$$

Where σ is an estimated scale parameter and where $f_m = \frac{1-\alpha_m}{{x_m}^* + \gamma_m}$.

Socio-demographic variables are included at the discrete part of the model through ψ_j.

5.3.1.4. Multiple discrete–continuous nested extreme value model (MDCNEV)

To accommodate the correlation among alternatives in the MCDEV model, Pinjari and Bhat (2010) introduced the nesting structure of MDCEV model. Their proposed nesting structure assumed that the unobserved part of the utility of the different activities followed a joint extreme value distribution The probability for the MDCNEV model can be written as follows:

$$P({x_1}^*, {x_2}^*, \dots\dots, {x_M}^*, 0, \dots, 0) = |J| \frac{\prod_{i \in chosen\ alts} e^{\frac{v_i}{\theta_i}}}{\prod_{s=1}^{S_M}\left(\sum_{i \in s^{th} nest} e^{\frac{v_i}{\theta_s}}\right)^{q_s}} \quad (12)$$

$$. \sum_{r_1=1}^{q_1} \cdots \sum_{r_s=1}^{q_s} \cdots \sum_{r_{S_M}=1}^{q_{S_M}} \left\{ \prod_{s=1}^{S_M} \left[\frac{\left(\sum_{i \in s^{th} nest} e^{\frac{V_i}{\theta_s}}\right)^{\theta_s}}{\sum_{s=1}^{S_j}\left\{\left(\sum_{i \in s^{th} nest} e^{\frac{V_i}{\theta_s}}\right)^{\theta_s}\right\}} \right]^{q_s - r_s + 1} \left(\prod_{s=1}^{S_M} \sum X_r s\right) \left(\sum_{s=1}^{S_M}(q_s - r_s + 1) - 1\right)! \right\}$$

.

Here, 1, 2, …, S_M is the nest that have M chosen alternatives. $q_1, q_2, \dots, q_{S_M}$ is the number of chosen alternatives in each of the S_M and $q_1 + q_2 + \cdots + q_{S_M} = M$.

5.3.1.5. Latent class MDCEV model (LC-MDCEV)

To capture the heterogeneity in satiation (non-satiation) using an MDCEV framework, it is assumed that respondents belong to B classes. The probability (on b) for the sequence of choices can be written as follows:

$$L_n(\beta) = \sum_{b=1}^{B} \pi_{n,b} \prod_{t=1}^{T_n} P_{j,n,t}(\beta_b) \tag{13}$$

Here, π_b is the probability of an individual belonging to class b, where $0 < \pi_b < 1$ and $\sum_{b=1}^{B} \pi_b = 1$ for each individual. The probability π_b of an individual belonging to class b can be calculated using logit formulation.

All the models estimated in this study used the "Apollo (0.2.9)" package R.

5.3.2. Modelling issues

To answer our research questions, we estimated multiple types of FMNL, FNL, MDCEV, MDCNEV, and Latent class MDCEV models. The base FMNL and MDCEV models were estimated by maintaining the base alternative basic needs as constant for both models (without covariates). Progressively different socio-demographic and social network variables were added with the utility. For ease of comparison, similar covariates were added with the fractional logit and multiple discrete continuous models. The majority of the variables were included in the model in the form of dummy variables. We tested the effects of age, gender, marital status, whether or not living with a partner, working status of respondents/partner (for last 6 months), income, education, access to a car, driving license status, access to ICT (internet, laptop, computer, smartphone), whether or not have underaged children in the household, dwelling condition, etc. and only the coefficients which were statistically significant were retained. Besides the socio-demographic factors, the effect of social network size (NALT), the proportion of social network living within the same neighbourhood (PVEC), and proportion of social network close to the respondent (PMC) factors were also tested. For the MDCEV and MDCNEV models, only baseline translation parameters were estimated for MDCEV and MDCNEV models. The baseline utility and satiation parameters are described in sections 5.4.1 and 5.4.2. To capture heterogeneity in satiation, LC-MDCEV models were estimated where both baseline and satiation parameters varied across different classes. Additionally, posterior analysis was carried out to identify the profile of each class. Parameter estimates from the LC-MDCEV models are summarised in section 5.4.3. Furthermore, to capture the correlation among the alternatives, all the potential nesting structures were tested and the best five are summarised in section 5.4.2.

5.4. Results

5.4.1. In-home and out-of-home activity participation

The results from the final specification of the FMNL and MDCEV models, including socio-demographic and social network variables, are presented in Table 5-3. The inclusion of socio-demographic variables with the base FMNL and MDCEV model significantly improved the model fit (results from the base FMNL and MDCEV are included in Supplementary Table 5-1. ASCs estimated from the base FMNL model (without covariates) illustrated the proportion of time respondents were likely to allocate to other available activities compared to the base alternative, i.e., *basic needs.* The relative values of ASCs from the base FMNL model aligned with the average time allocation by the respondents (Table 5-2). For instance, respondents tend to spend a higher proportion of their time on out-of-home work activities, followed by household obligations, and in-home leisure activities, respectively (Supplementary Table 5-1). Conversely, baseline utility constants from the base MDCEV model reflected how frequently the sample chose each activity, consistent with discrete choices (Table 5-2). For example, the results from the baseline constants of MDCEV indicated that respondents were more likely to choose travel, in-home leisure, and out-of-home social activities, respectively (Supplementary Table 5-1). However, the negative signs of all ASCs corresponding to inside goods in MDCEV suggested that all individuals in our sample allocate some non-zero time to meet basic needs (the essential outside good). Like the base FMNL and MDCEV models, the estimates from the final specification of FMNL and MDCEV have different implications. The ASCs of the FMNL model reflected the average marginal effect on the share of time allocation for the reference group, such as individuals aged over 50, females, unmarried, those without a driving license, partners not working for the last 6 months, households without underaged children, and those lacking access to ICT. On the contrary, the baseline constants of MDCEV for the final specification indicated the likelihood of the reference group choosing a particular activity. Hence, the covariates that interacted with the FMNL model reflected their average marginal effect on time allocation, while in the MDCEV, they signified respondents' sensitivity in choosing a particular activity. The results from both the FMNL and MDCEV models, along with the significance of various socio-demographic factors affecting the utility of time proportion and choice of activity type, aligned intuitively with existing studies conducted within a similar cultural context.

The results from the final FMNL model indicated that individuals under 22 years of age were likely to allocate a higher proportion of their time to education and social activities, whereas decision-makers between 22 and 50 years old were more inclined to dedicate a higher share of their time to household obligations and work activities compared to the reference group (those aged over 50 years old). The baseline parameter estimates from MDCEV confirmed that individuals under 22 years old were more likely to prioritise school and social activities, while those aged between 22 and 50 were more inclined to focus on household and work activities compared to the reference group. Moreover, gender was identified as a factor influencing household obligations, with men tending to allocate a smaller share of their time to this activity, as indicated by the FMNL results, suggesting a lower inclination to participate in them according to the MDCEV findings. Marital status also demonstrated significance both in the FMNL and MDCEV models, impacting respondents' decisions regarding the allocation of time and likelihood of participation in personal and shopping activities. The influence of holding a driving license on accompanying someone appeared intuitive across both the FMNL and MDCEV models, with respondents showing a higher propensity to invest time in accompanying activities and a greater likelihood of participation, respectively compared to those without a driving license. The results regarding age, gender, marital status, and effect of driving license aligned with a similar study conducted by Calastri et al. (2017).

The influence of partners working for the last 6 months highlighted that respondents were likely to participate in in-home work activities compared to those whose partner was not working for the last 6 months. The descriptive statistics highlight that the majority of the respondents whose partners have been working for the last 6 months were also working for the last six months (approximately 66%), are female (approximately 70%) and have underaged children (65%). Higher household obligations and responsibilities may have compelled them to work from home or invest some time at home after office hours. Furthermore, given their age distribution, which suggested that they were predominantly in the working stage of life, it is intuitive that they invested less time in in-home education activities. Also, the result revealed that respondents with underaged children in the household were likely to spend less proportion of time for in-home leisure and more time for in-home social time compared to others without underaged children. Parents may find themselves dedicating more time to caring for their children and engaging in activities such as feeding, playing, or supervising them, which leaves less time for personal leisure pursuits. However, despite the potential reduction in leisure time, there may be an increase in in-home social activities. This could be because parents may prioritise spending quality time with their children and engaging in family-oriented social activities within the home environment. These activities may include family meals, games, or other interactive pursuits that involve both parents and children. Additionally, parents might invite other families with children over for social gatherings or playdates, leading to increased in-home social interactions. This result is supported by the effect of the proportion of social network members residing in the respondent's neighbourhood and the size of their social network. With the increase in the proportion of social network living in the respondents' neighbourhood results from FMNL showed a lower share of time investment for out-of-home leisure and social activities which was also highlighted in the baseline preference parameters of MDCEV showing less likelihood of engagement on those activities. Similarly, with the increase of social network size respondents were likely to invest a higher share of their time to in-home social activities as they were likely to engage in those activities. The influence of access to ICT was also tested in determining its effect on the proportion of time investment in different alternatives in FMNL and the likelihood of activity participation in the MDCEV model. The result appeared plausible as access to ICT showed a higher propensity to invest a higher share of time in in-home education and in-home leisure compared to those without access to ICT. Though the significance of ICT's influence on choosing in-home leisure activities was insignificant, however, was significant for in-home education activity.

An additional outcome obtained from the MDCEV model was the satiation parameters, which reflected respondents' sensitivity to less rapid or more rapid satiation in activity participation. Since this study aimed to compare the prediction accuracy of activity duration either considering the diminishing marginal utility or without considering it, we didn't further parameterise the translation parameters to accommodate heterogeneity in satiation across decision-makers. Only baseline satiation was estimated to take into account the satiation in activity duration prediction. The coefficients shown in Table 5-3 highlighted that respondents were less rapidly satiated by both out-of-home and in-home work and education activities compared to the other activities.

Eventually, we compared the time-use prediction of the FMNL and MDCEV models with the observed time use and measured the root mean square error (RMSE) to check their corresponding accuracy since these two models could not be compared using the goodness of fit measures. The prediction summary of FMNL and MDCEV is represented in Table 5-4 (with covariates) and Supplementary Table 5-2 (without covariates). Overall, the results from the FMNL model reinforced that the utility of activity duration with implicit satiation effect (captured by non-linearity) can easily recover the expected average values of the dependent variable (i.e., the time use) compared to the MDCEV model. Further, the time use prediction using the FMNL yielded slightly better prediction than the MDCEV model (RMSE in the FMNL and MDCEV model were respectively 1.87 and 1.89). RMSE was approximately the same for both the FMNL and MDCEV models for several activities, such as in-home work, education, and social activities, as well as out-of-home education, and personal activities, where participation frequency was relatively lower than the other activities. In contrast, the FMNL model had relatively lower RMSE than the MDCEV model for certain in-home and out-of-home activities (basic needs, household obligation, in-home leisure, work, shop, accompany/Drop-off and travel), where the observed percentage of engagement rate was higher (Table 5-2 and Table 5-4). The exception was observed for the out-of-home social activity, where the proportion of observed participation rate was higher but the RMSE from FMNL and MDCEV was approximately the same. Similarly, the FMNL showed improved prediction accuracy when the observed share (%) of accompany/drop-off engagement was lower compared to other activities. It is worth noting that the relative time spent in the accompany/drop-off activity was the lowest among the other activities if participated.

Table 5-3 Estimates from the comprehensive FMNL and MDCEV model.

	FMNL		**MDCEV**	
	Estimate	**R.t.rat.(0)**	**Estimate**	**R..rat.(0)**
Utility constants (δ)				
Basic needs	0	NA	0	NA
Household obligation	-2.032	-14.635	-2.404	-16.830
In-home work	-3.880	-11.963	-4.799	-19.117
In-home education	-4.792	-7.776	-5.843	-12.138
In-home leisure	-1.854	-17.649	-2.303	-19.720
In-home social	-4.329	-20.913	-4.569	-23.327
Work	-1.729	-9.831	-3.159	-17.980
School	-4.338	-13.185	-5.511	-18.297
Personal issues	-3.853	-15.549	-4.089	-22.390
Shop	-3.586	-27.927	-3.124	-24.519
Leisure	-3.216	-20.829	-4.098	-29.007
Social	-1.929	-22.856	-2.464	-29.028
Accompany/Drop-off	-4.539	-20.147	-4.617	-20.629
Travel	-2.048	-42.168	-0.865	-7.820

Satiation parameters (φ)					
Household obligation				1.590	14.578
In-home work				2.443	6.972
In-home education				3.544	6.348
In-home leisure				1.611	17.082
In-home social				1.482	13.89
Work				6.334	13.952
School				4.100	5.273
Personal issues				1.133	12.182
Shop				0.469	11.495
Leisure				2.319	11.089
Social				1.410	15.429
Accompany/Drop-off				0.622	10.723
Travel				0.250	10.635
Socio-demographic variables					
Age <22	School	2.637	5.491	2.679	6.070
	Social	0.672	3.464	0.837	3.719
Age 22-50	Household obligation	0.527	3.320	0.294	1.839
	Work	0.480	2.428	0.363	1.841
Gender (Male)	Household obligation	-0.804	-3.377	-0.964	-5.184
	Travel	0.146	1.970	0.282	3.603
Marital status (Married)	Personal issues	0.649	2.163	0.489	2.342
	Shop	0.466	2.942	0.294	2.056
Having driving license	Accompany/Drop -off	0.652	2.308	1.152	4.372
Partner working for last 6 months	In-home work	0.732	1.753	0.643	2.029
	In-home education	-2.094	-2.640	-1.937	-2.675
Have underaged children in the household	In-home leisure	-0.398	-2.593	-0.457	-3.114
	In-home social	0.582	2.401	0.300	1.393
Access to ICT	In-home education	1.802	2.549	1.749	2.995
	In-home leisure	0.335	2.074	0.166	1.112
PVEC*	Leisure	-0.856	-1.785	-0.577	-1.096
	Social	-0.462	-1.723	-0.505	-1.904

NALT	In-home social	0.036	3.690	0.026	2.774
Goodness of fit measures					
LL (Start)			-3132.56		-23902.74
LL (Final)			-2117.48		-17888.63
AIC			4296.96		35865.27
BIC					36088.75

*Proportion of social network are living in respondent's neighbourhood

Table 5-4 Prediction summary of FMNL and MDCEV model (with covariates).

		FMNL	MDCEV	FMNL	MDCEV
	Observed time-use	**Predicted time-use**	**Predicted time-use**	**RMSE**	**RMSE**
Basic needs	11.68	11.68	11.34	2.97	2.98
Household obligation	1.77	1.77	1.75	2.29	2.30
In-home work	0.37	0.37	0.34	1.33	1.33
In-home education	0.15	0.15	0.15	0.81	0.81
In-home leisure	1.74	1.74	1.81	2.22	2.24
In-home social	0.29	0.29	0.33	0.92	0.92
Work	2.96	2.96	2.47	3.69	3.74
School	0.31	0.31	0.30	1.28	1.28
Personal issues	0.40	0.40	0.43	1.17	1.17
Shop	0.44	0.44	0.51	0.90	0.91
Leisure	0.41	0.41	0.41	1.26	1.26
Social	1.69	1.69	1.66	2.43	2.43
Accompany/Drop-off	0.19	0.19	0.26	0.53	0.54
Travel	1.60	1.60	2.25	1.31	1.48
Overall				**1.87**	**1.89**

5.4.2. Correlation among the alternatives

To investigate the potential impact of unobserved correlation (among the alternatives) on findings related to satiation in predicting activity duration, we tested different nesting structures using both fractional logit model and MDCEV model. The goodness of fit measures of the MDCNEV model showed a clear improvement in the model fit while capturing the correlation among the alternatives (Table 5-5), however, the FNL model fit did not (Supplementary Table 5-5). Moreover, the improvement in the model fit varied depending on the nest structure used to accommodate the correlation. In Table 5-5, different specifications are ordered by their goodness of fit which was measured by the final LL value. We also included the AIC and BIC values used to compare the models with each other. It is apparent in Table 5-5 that MDCNEV 1 had the best fit (LL and BIC) while capturing the correlation among alternative activities, where different levels of obligation can be associated with varying degrees of involvement (LL: -17456.66 and BIC: 35238.97). Despite better model fits of all five reported MDCNEV models, their RMSE values (calculated based on predicted time use) were very similar to the MDCEV model for most of the nesting structures and, for different alternatives (Table 5-6). Although the MDCNEV 1 model, which exhibited the best model fit, had a slightly higher overall RMSE compared to the MDCEV model, there were instances where the RMSE for specific individual activities—such as household obligations and in-home work—was approximately the same. On the other hand, the RMSE of work activities was found to be lower in the MDCNEV 1 model compared to the MDCEV model. Furthermore, significant decreases in RMSE values were noted for certain activities in the MDCNEV 1 model when errors were calculated considering both the discrete and continuous predictions. Supplementary Table 5-3 shows that the MDCNEV 1 model had lower RMSE values for time-use prediction (if participated) for the majority of activities, with the exception of in-home leisure, social, and travel activities. Compared to the other nesting structure, the best fit MDCENV model showed lower RMSE for time use prediction (if participated).

Besides the model fit and RMSE of the estimated models, we compared the prediction accuracy of both the discrete and continuous components (considering satiation) of MDCEV and MDCNEV models. The summary of predicted time use is reported in Table 5-7 and Table 5-8, and the predicted discrete probability distribution in Supplementary Figure 5-1. While comparing the discrete probability (Supplementary Figure 5-1), the best fit MDCNEV model showed better agreement with the observed mean probability. However, when MDCNEV 1 model improved the discrete probability, it encountered challenges in predicting continuous components i.e. time use (with satiation). Table 5-7 shows the overestimation for certain activities such as for in-home leisure, in-home social, personal issues, shopping, leisure, social, accompanying, and travel, or underestimation for basic needs and household activities compared to the MDCEV model. However, the predicted time use (if participated) considering both discrete and continuous component of MDCNEV 1 exhibited a relatively better agreement with the observed average time use (if participated) compared to the other MDCNEV and MDCEV model. Eventually, the higher deviation from the observed average time use (if participated) was found for certain activities (e.g., basic needs, in-home and out-of-home leisure, and social activities) which were grouped to capture correlation among the other alternatives and nest parameters were significantly different from 1 (insinuating unobserved significant correlation among the alternatives). A similar effect was also observed in the other MDCNEV model i.e., having a strong correlation among the alternatives but a relatively higher deviation of predicted time use compared to the observed time use (if participated).

Table 5-5 Different nesting specification and MDCNEV model fit.

		Nest 1	Nest 2	Nest 3	LL (final)	AIC	BIC
1	Nest 1 (Obligation involved): Basic needs, household obligation, in-home work, in-home education, work, school Nest 2 (Less obligation involved): In-home leisure, in-home social, personal issues, shop, leisure, social, accompany/drop-off	(0.266) ***	(0.822) ***		-17456.66	35005.33	35238.97
2	Nest 1 (In home activities): Household obligation, in-home work, in-home education, in-home leisure, in-home social Nest 2 (Out-of-home activities): Work, school, personal issues, shop, leisure, social, accompany/drop-off, travel	(0.873) ***	(0.440) ***		-17579.39	35250.78	35484.43
3	Nest 1 (Mandatory work): In-home work, Work Nest 2 (Discretionary activities): In-home leisure, in-home social, leisure, social, shop, personal issues, accompany/drop-off	0.967 ***	(0.658) ***		-17791.21	35674.41	35908.06
4	Nest 1: In-home social, social, travel Nest 2: In-home work, work Nest 3: Household obligation, in-home education, in-home leisure, school, personal, shop, leisure, accompany/drop-off	(0.661) ***	0.945***	(0.875) ***	-17799.78	35693.56	35932.28
5	Nest 1: Personal, shop, accompany/drop-off, travel Nest 2: In-home social, social	(0.710) ***	(1) ***		-17829.62	35751.23	35984.87

*** All the values are significant at 95% confidence level, ** All the values are significant at 90% confidence level, * All the values are significant at 80% confidence level, (…) at 95% confidence level values were significantly different from 1

Table 5-6 RMSE of the MDCEV and different MDCNEV models.

	MDCEV	MDCNEV 1	MDCNEV 2	MDCNEV 3	MDCNEV 4	MDCNEV 5
Basic needs	2.98	3.24	3.01	2.99	3.00	3.00
Household obligation	2.30	2.30	2.29	2.30	2.29	2.30
In-home work	1.33	1.33	1.33	1.33	1.33	1.33
In-home education	0.81	0.81	0.81	0.81	0.81	0.81
In-home leisure	2.24	2.26	2.24	2.24	2.24	2.24
In-home social	0.92	0.92	0.92	0.92	0.93	0.92
Work	3.74	3.67	3.69	3.73	3.74	3.73
School	1.28	1.28	1.28	1.28	1.28	1.28
Personal issues	1.17	1.17	1.17	1.17	1.17	1.18
Shop	0.91	0.91	0.90	0.91	0.91	0.91
Leisure	1.26	1.27	1.27	1.27	1.26	1.26
Social	2.43	2.43	2.43	2.43	2.43	2.42
Accompany/Drop-off	0.54	0.55	0.53	0.54	0.54	0.55
Travel	1.48	1.52	1.32	1.36	1.40	1.42
Overall	1.89	1.92	1.88	1.89	1.89	1.89

Table 5-7 Observed and predicted time use summary of different MDCNEV models.

	Observed	MDCEV	MDCNEV 1	MDCNEV 2	MDCNEV 3	MDCNEV 4	MDCNEV 5
Basic needs	11.68	11.34	10.35	11.63	11.45	11.28	11.31
Household obligation	1.77	1.75	1.55	1.81	1.80	1.74	1.75
In-home work	0.37	0.34	0.34	0.38	0.36	0.36	0.34
In-home education	0.15	0.15	0.15	0.17	0.15	0.16	0.14
In-home leisure	1.74	1.81	2.05	1.86	1.80	1.79	1.80
In-home social	0.29	0.33	0.40	0.36	0.37	0.45	0.31
Work	2.96	2.47	2.71	2.81	2.50	2.45	2.45
School	0.31	0.30	0.32	0.37	0.30	0.33	0.29
Personal issues	0.40	0.43	0.51	0.43	0.46	0.44	0.55
Shop	0.44	0.51	0.58	0.43	0.50	0.50	0.58
Leisure	0.41	0.41	0.51	0.47	0.48	0.43	0.41
Social	1.69	1.66	1.90	1.55	1.67	1.80	1.64
Accompany/Drop-off	0.19	0.26	0.30	0.22	0.26	0.26	0.32
Travel	1.60	2.25	2.34	1.52	1.89	2.03	2.10

Table 5-8 Observed and predicted time use summary (if participated) of different MDCNEV models.

	Observed	MDCEV	MDCNEV 1	MDCNEV 2	MDCNEV 3	MDCNEV 4	MDCNEV 5
Basic needs	11.68	11.34	**10.35**	11.63	11.45	11.28	11.31
Household obligation	3.22	4.32	3.44	4.31	4.35	4.26	4.33
In-home work	3.11	4.41	3.29	4.57	4.52	4.55	4.45
In-home education	3.52	5.34	3.86	5.30	5.17	5.31	5.14
In-home leisure	3.02	4.36	**4.54**	4.39	4.29	4.32	4.37
In-home social	1.94	3.35	**3.25**	3.41	3.11	3.61	3.29
Work	6.59	7.74	7.10	7.97	7.76	7.75	7.73
School	4.40	5.69	4.65	5.46	5.64	5.79	5.66
Personal issues	1.83	2.93	2.84	2.46	2.72	2.94	3.14
Shop	1.09	1.81	1.78	1.46	1.64	1.77	1.93
Leisure	2.76	4.32	**4.22**	3.86	4.14	4.40	4.35
Social	2.97	4.04	4.17	3.81	3.97	4.25	4.02
Accompany/Drop-off	0.96	1.96	1.88	1.44	1.97	1.91	2.07
Travel	1.74	2.72	2.80	1.89	2.33	2.49	2.56

5.4.3. Heterogeneity in satiation

To further demonstrate the significance of random heterogeneity in activity participation and time use, we estimated latent class MDCEV models where parameters (satiation and baseline) varied across different classes. The rationale for relaxing the homogeneity in the baseline and satiation assumption was the model prediction accuracy's dependency on the observed activity participation share and corresponding time usage (both overall average time use, and time use if participated). We started from class size 2 and gradually increased the class size by observing the resulting classes' corresponding fit measures and interpretability. The LL values of the latent-class models (two-class, three-class, and four-class) were compared with each other and the MDCEV model (includes sociodemographic) respectively. As shown in Table 5-9, the LL showed an improvement as the number of classes increased. In models where both baseline parameters and satiation parameters were allowed to vary across different classes, a log-likelihood ratio (LR) test indicated significant improvement in LL (95% confidence level) up to 3 classes, with improvement in both the AIC and BIC values. Further, to distinguish the contributions of baseline and satiation parameters to the improvement in LL, we estimated separate latent class models by allowing either the baseline or satiation parameters to vary across different classes. Results revealed that the significance of the improvement in model fit relied on which parameters were allowed to vary across different classes. Heterogeneity in baseline parameters across different classes demonstrated a significant contribution to the improvement of the model fit compared to the satiation parameters. Furthermore, we also conducted t-tests to investigate whether parameters (baseline and satiation) from each class for each activity differed significantly from those of other classes. The t-test results also highlighted significant variation in heterogeneity across different classes, with higher variations observed in baseline parameters compared to the satiation parameters (Supplementary Table 5-4).

Table 5-10 presents the three class MDCEV model where both satiation and baseline parameters varied across different classes. The result from the three class MDCEV model revealed that apart from the socio-demographic variables considered in the baseline utility, there was significant unobserved heterogeneity across different classes. This variation was more discernible across different classes while comparing the baseline parameters of both in-home and out-of-home activities compared to the satiation parameters (Supplementary Table 5-4). Level of satiation varied across different classes significantly for work activities, personal, and social activities. Moreover, to understand the characteristics of each latent class we carried out a posterior analysis which is summarised in Table 5-11. The posterior analysis showed that among the unobserved subgroups, class 1 (29.6%) was dominated by the respondents who were female, married or living with a partner and had a partner working for the last 6 months. Compared to the other two sub-groups, this particular group exhibited a greater inclination towards engaging in in-home activities as well as out-of-home personal activities and shopping. Regarding the satiation parameters, the likelihood of engaging in educational activities showed less rapid satiation, although not significantly different from the other groups. However, for personal activities and shopping, there was a higher likelihood of opting for longer activity hours, which was significantly different from subgroup class 2. On the contrary, class 2 (48.9%), comprising the largest share of the sample, was predominantly characterised by the individuals aged between 22 to 50 years, having been employed for the last 6 months, and having a car at the household. The baseline parameters aligned with the expectation (as dominated by the working age group) that this group exhibited a higher propensity for work-related activities. Intuitively, the satiation parameters indicated rapid satiation for personal, shopping, social, and accompanying activities, with their level of satiation significantly differing from that of class 2. On the other hand, Class 3 (21.5%) was dominated by the respondents who were aged below 22 years old and students, who were less likely to participate in in-home (obligatory and discretionary) activities, but more likely to participate in out-of-home discretionary activities. Also, the satiation parameters of class 3 indicated that they were likely to be less rapidly satiated for out-of-home social activities.

It should be noted that a similar approach to capturing unobserved heterogeneity across different classes for time use decisions was explored using the time use model by FMNL. However, the latent class model did not reject the one-class model. Therefore, we only reported the investigation based on the latent class MDCEV model.

Table 5-9 Latent class MDCEV model description and model summary.

Model ID	**Model description**	**Goodness of fit measures**	
1	Base MDCEV without socio-demographics	N. parameters	26
		Log-likelihood	-18144.53
		AIC	36341.05
		BIC	36473.11

2	MDCEV with socio-demographics	N. parameters	44		
		Log-likelihood	-17888.63		
		AIC	35865.27		
		BIC	36088.75		
			Both baseline and satiation parameters vary across classes	***Only satiation parameters vary across classes***	***Only baseline parameters vary across classes***
3	Latent class MDCEV model with two classes	N. parameters	60	53	52
		Log-likelihood	-17679.44	-17879.78	-17698.90
		AIC	35478.88	35865.56	35501.81
		BIC	35783.63	36134.76	35765.92
4	Latent class MDCEV model with three classes	N. parameters	78	62	60
		Log-likelihood	-17537.61	-17875.73	-17561.21
		AIC	35231.21	35875.46	35242.42
		BIC	35627.39	36190.37	35547.17
5	Latent class MDCEV model with four classes	N. parameters	92	-	68
		Log-likelihood	-17514.15		-17501.99
		AIC	35212.29		35139.97
		BIC	35679.58		35485.36

Table 5-10 Latent class MDCEV model result (baseline preference and satiation parameter estimates.

	Class 1 29.6%	Class 2 48.9%	Class 3 21.5%
Baseline parameters			
Household obligation	-1.480***	-2.352***	-4.187***
In-home work		-4.821***	
In-home education		-5.852***	
In-home leisure	-1.588***	-2.131***	-3.072***
In-home social	-4.042***	-4.749***	-5.050***
Work	-4.699***	-2.219***	-4.078***
School		-5.494***	
Personal issues	-3.623***	-4.378***	-4.231***
Shop	-3.050***	-3.080***	-3.358***
Leisure	-4.307***	-4.586***	-3.554***
Social	-2.755***	-2.380***	-2.376***
Accompany/Drop-off		-4.604***	
Travel		-0.847***	

Satiation parameters				
Household obligation		1.081***	1.248***	1.443***
In-home work			2.454***	
In-home education			3.519***	
In-home leisure		1.444***	1.393***	1.959***
In-home social			1.476***	
Work		3.550***	3.870***	7.824***
School		5.700**	2.606***	4.130***
Personal issues		1.553***	0.823***	1.030***
Shop		0.630***	0.376***	0.549***
Leisure			2.238***	
Social		1.602***	1.133***	1.801***
Accompany/Drop-off		0.755***	0.534***	0.714***
Travel			0.247***	
Age <22	School			2.486***
	Social			0.596***
Age 22-50	Household obligation			0.467**
	Work			0.322*
Gender (Male)	Household obligation			-1.151***
	Travel			0.286***
Marital status	Personal issues			0.503***
	Shop			0.301***
Having driving license	Accompany/Drop-off			1.118***
Partner working for last 6 months	In-home work			0.687***
	In-home education			-1.902***
Have underaged children in the household	In-home leisure			-0.731***
	In-home social			0.211*
Access to ICT	In-home education			1.726*
	In-home leisure			0.194***
PVEC	Leisure			-0.097*
	Social			-0.296*

NALT	In-home social	0.027^{***}

Table 5-11 Class definition by socio-demographic variables.

	Class 1	Class 2	Class 3
	29.6%	48.9%	21.5%
Age below 22	6.12%	2.42%	26.78%
Age between 22 to 50	67.91%	72.56	69.84%
Female	70.12%	54.83	67%
Currently student	4.07%	11.32%	44.84%
Married or living with partner	78.47%	74.65%	42.70%
Working for last 6 months	36.63%	90.87%	44.46%
Partner working for last 6 months	64.58%	49.50%	36.70%
Have car at the household	69.61%	72.62%	59.26%
Have driving license	48.70%	57.72%	58.84%
Have access to ICT	42.35%	33.66%	42.86%
Sharing living space with other	30.53%	25.37%	36.93%
Have access to domestic service	8.15%	11.18%	31.48%

5.5. Discussion

This study investigates the influence of satiation variation on activity type and time use modelling. To address research question 1, which focused on the effect of satiation in predicting activity duration, we conducted a comparison of model fit and prediction accuracy between the FMNL and MDCEV models. This comparison specifically targeted in-home and out-of-home activity types and time use decisions. Furthermore, we investigated whether dealing with correlation affects the findings in terms of satiation. Additionally, we evaluated the ability of the latent class MDCEV model to capture the effects of unobserved heterogeneity, with both baseline and satiation parameters varying across different classes. Our analysis utilised a weekly time use dataset from Concepción, Chile. The comparative analysis provided insights into the performance and prediction accuracy of activity duration models. The FMNL model exhibited better agreement with observed average time use. Additionally, the MDCNEV model achieved a better fit by capturing observed correlations among the alternatives. While the MDCNEV's prediction of activity duration did not show a significant reduction in RMSE when measured without considering discrete choice probability, the model consistently and accurately estimated average activity duration when both discrete and continuous predictions were employed. After accommodating correlations among the alternative activities, significant improvements were observed in the MDCNEV models for discrete prediction, albeit at the expense of predicting average activity duration. On the other hand, the latent class MDCEV model demonstrated improved model fit by relaxing the assumption of homogeneity in baseline and satiation parameters. Within the latent class MDCEV model, the effect of heterogeneity in baseline parameters outperformed the effect of unobserved heterogeneity across different classes.

Results from the FMNL, MDCEV, and MDCNEV models showed that the activity type and patterns (time commitment if engaged in the activity), correlation among the alternatives, as well as the specification of duration (explicit or implicitly accommodating satiation) in the model, affected the prediction accuracy of time use. Activity duration of activities performed by the major proportion of the sample can be predicted more parsimoniously and accurately by the FMNL model. This is because FMNL is a model of the mean of the dependent variable conditional on the covariates (Murteira and Ramalho, 2016). On the other hand, the explicit accommodation of the satiation component in the MDCEV model did not enhance the prediction accuracy of duration for activities performed by the higher and lower proportion of the sample. Similarly, the FMNL model anticipated more accurately the average amount of time spent on both short and long duration activities. Therefore, activities that are often participated in and demand less time commitment can predict activity duration with higher accuracy using the FMNL model. Though MDCEV is flexible enough to forecast both discrete and continuous components of activity choice decisions, it falls short of recovering the average time usage, which is a strength of FMNL. Furthermore, to ensure unbiased parameter estimates and accurate predictions, it is crucial to account for correlations among alternatives. This allows for a more precise estimation of the relationships between variables and the decision-making process (Pinjari and Bhat, 2010). Moreover, Murteira and Ramalho (2016) highlighted that the IIA assumptions may not be consistent in the fractional response and Dubin (2007) suggested to use nesting framework to capture unobserved correlation among the alternatives. Hence, we tested different nesting frameworks to capture the unobserved correlations among the alternatives using the FMNL model (including both the constants and socio-demographics). Results obtained from the FNL model indicated that considering nesting parameters did not significantly improve the model fit. Therefore, in the fractional model, the nesting framework did not capture the unobserved correlation among the alternatives and improve the prediction accuracy of activity duration. Conversely, the MDCNEV model exhibited a substantial association among the alternatives and an enhanced model fit. However, despite the improved model fit with the MDCNEV model, there was not a corresponding improvement in the predicted time use accuracy. While MDCNEV improved the prediction of mean discrete probability, the prediction of average activity duration was further exacerbated when accounting for correlations among the alternatives. Improvement in time use prediction was observed while considering both the discrete and continuous component. As a result, the MDCEV framework demonstrated its strength in its predictive capacity for both participants and non-participants.

Moreover, results from the latent class MDCEV model indicated that the satiation effect varied across different classes based on the type of activity under consideration. Also, the model fit of the latent class model (where both baseline and satiation varied across different classes) suggested that the latent class MDCEV model had a better model fit compared to the MDCNEV model. This supports the assumption of heterogeneity in activity participation and time use across different classes while assuming homogeneity within each class. However, compared to heterogeneity in satiation, variability in baseline parameters across different classes has a more pronounced impact on improving the model fit. These findings highlight the need to investigate whether individuals are more inclined towards seeking variety rather than experiencing diminishing marginal utility.

Besides, the demographic profile of each class showed how satiation may differ based on the related sample profile. Additionally, the class definition and associated satiation effects demonstrated how contextual variables such as past experience, current working status, and access to certain facilities (e.g., access to car) may generate different likelihood of activity participation and satiation for different activities (e.g., individuals who have been employed for the past six months were more likely to participate in work-related activities and devote longer duration to such tasks; those with partners who have been employed for the past six months were more prone to engaging in in-home activities as well as discretionary activities outside the home). These results were also supported by similar studies related to capturing the heterogeneity in activity participation (Sobhani et al., 2013; Wafa et al., 2015; Astroza et al., 2017). Therefore, depending on the context, and sample profile, for some activities and demographic groups, satiation consideration may be useful, but it may not be worthwhile for other groups and activities. Homogeneity in satiation for the whole population segment may lead to either underestimating or overestimating the influence of satiation that characterise individuals' activity-travel decisions.

5.6. Conclusion

The aim of this study was to investigate the influence of satiation in predicting activity duration and evaluate its ability to capture variations in activity participation and satiation across different demographic groups, while also observing correlations among different alternatives. In this study, a multi-day travel and activity record was used to analyse the activity type and duration choice behaviour. Both discrete and continuous activity dimensions were simultaneously used for activity choice modelling. In response to the first research question which aimed to investigate whether incorporating satiation improves the prediction accuracy of activity duration showed that the FMNL model (implicitly accommodated satiation) performed better than the MDCEV model. Moreover, for our specific sample where the majority of the respondents were full-time workers or students, the prediction of mean time use using logit specification (FMNL model) was able to accurately predict the average time people spent on different activities. When addressing the part of our second research question, which sought to investigate whether dealing with correlation affects the prediction of duration while considering satiation. An improvement in the model fit was observed when different nesting structures were tested within the MDCNEV framework. On the contrary, when a similar nesting structure was applied to the FNL model, it was challenging to clearly observe this correlation effect. Although the RMSE indicated a relatively higher error in the best fitting MDCNEV model, the prediction of average time use (if participated) demonstrated better alignment with observed average time use. This improvement is primarily attributed to better prediction of discrete probability rather than continuous time use prediction with satiation.

We also observed a potential improvement in model fit with the MDCEV model when accounting for variations in activity participation and satiation. These findings signify the fact of varying degrees of influence of baseline parameters and satiation in improving model fit. In the selected sample, varying activity participation was more dominant compared to varying satiation across different classes. Moreover, with the given sample profile, the satiation effect varied based on their likelihood of choosing a particular activity, context, and sample characteristics. While Class 1 showed a less rapid satiation for most of the in-home activities, Class 2 exhibited less rapid satiation for work and Class 3 highlighted more inclination for longer time investment in non-home-based discretionary activities. Class size indicated an unequal distribution of observation between the identified groups where one group was more prevalent than the other group. Future research is required to examine the role of satiation using different sample profiles and activity patterns (different observation and time usage compositions) using the similar comparative analysis of other modelling approach. However, when comparing the MDCNEV model with the heterogeneity test, we observed a better model fit when accounting for the unobserved correlations among the alternatives. Therefore, a latent class MDCNEV model can also be estimated in the future to determine the joint contribution of heterogeneity and correlation to the overall model fit and prediction accuracy. Findings from this study advise performing a heterogeneity test and comparative scenario analysis to identify the appropriate modelling framework for activity modelling and predicting activity duration. In comparison to a more complex model, choosing a suitable modelling framework that accurately depicts the underlying structure of the data might be helpful in terms of model interpretability and parsimony.

5.7. References

Astroza, S., Pinjari, A.R., Bhat, C.R. and Jara-Díaz, S.R. 2017. A Microeconomic Theory–Based Latent Class Multiple Discrete–Continuous Choice Model of Time Use and Goods Consumption. *Transportation Research Record.* **2664**(1), pp.31-41.

Axhausen, K.W. 2007. Definition of movement and activity for transport modelling. *Handbook of transport modelling.* Emerald Group Publishing Limited, pp.329-343.

Axhausen, K.W., Horni, A. and Nagel, K. 2016. *The multi-agent transport simulation MATSim.* Ubiquity Press.

Becker, G. 2014. *The portfolio structure of German households: A multinomial fractional response approach with unobserved heterogeneity.* University of Tübingen Working Papers in Economics and Finance.

Bernardo, C., Paleti, R., Hoklas, M. and Bhat, C. 2015. An empirical investigation into the time-use and activity patterns of dual-earner couples with and without young children. *Transportation Research Part A: Policy and Practice.* **76**, pp.71-91.

Bhat, C.R. 2005. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological.* **39**(8), pp.679-707.

Bhat, C.R. 2008. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological.* **42**(3), pp.274-303.

Bhat, C.R. and Srinivasan, S. 2005. A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B: Methodological.* **39**(3), pp.255-278.

Bhat, C.R., Srinivasan, S. and Sen, S. 2006. A joint model for the perfect and imperfect substitute goods case: application to activity time-use decisions. *Transportation Research Part B: Methodological.* **40**(10), pp.827-850.

Calastri, C., Hess, S., Daly, A. and Carrasco, J.A. 2017. Does the social context help with understanding and predicting the choice of activity type and duration? An application of the Multiple Discrete-Continuous Nested Extreme Value model to activity diary data. *Transportation Research Part A: Policy and Practice.* **104**, pp.1-20.

Cardoso, A.R., Fontainha, E. and Monfardini, C. 2010. Children's and parents' time use: empirical evidence on investment in human capital in France, Germany and Italy. *Review of Economics of the Household.* **8**, pp.479-504.

Carrasco, J.A., Hogan, B., Wellman, B. and Miller, E.J. 2008. Collecting social network data to study social activity-travel behavior: an egocentric approach. *Environment and Planning B: Planning and Design.* **35**(6), pp.961-980.

Chen, C.-C. and Liao, Y.-S. 2019. From satisfaction to satiation: The gradation model of satiation in casual leisure. *Annals of Leisure Research.* **22**(2), pp.233-246.

Dubin, J.A. 2007. Valuing intangible assets with a nested logit market share model. *Journal of econometrics.* **139**(2), pp.285-302.

Eom, J.K., Lee, K.-S. and Seong, M.-E. 2020. Development and application of the Activity-BAsed Traveler Analyzer (ABATA) system. *Future Generation Computer Systems.* **106**, pp.135-153.

Galak, J., Kruger, J. and Loewenstein, G. 2011. Is variety the spice of life? It all depends on the rate of consumption. *Judgment and Decision making.* **6**(3), pp.230-238.

Golshani, N., Shabanpour, R., Auld, J. and Mohammadian, A. 2018. Activity start time and duration: incorporating regret theory into joint discrete–continuous models. *Transportmetrica A: transport science.* **14**(9), pp.809-827.

Habib, K.M.N. 2012. Modeling commuting mode choice jointly with work start time and work duration. *Transportation research part A: Policy and practice.* **46**(1), pp.33-47.

Hamad, K. and Obaid, L. 2022. Tour-based travel demand forecasting model for a university campus. *Transport Policy.*

Hensher, D.A., Rose, J.M. and Greene, W.H. 2008. Combining RP and SP data: biases in using the nested logit 'trick'–contrasts with flexible mixed logit incorporating panel and scale effects. *Journal of Transport Geography.* **16**(2), pp.126-133.

Kapur, A. and Bhat, C.R. 2007. Modeling adults' weekend day-time use by activity purpose and accompaniment arrangement. *Transportation Research Record.* **2021**(1), pp.18-27.

Kochaniak, K. 2020. How does households' wealth affect portfolio choices?

Koppelman, F.S. and Wen, C.-H. 1998. Nested Logit Models Which Are You Using? *Transportation Research Record.* **1645**(1), pp.1-7.

Lee, L.-F. 1982. Some approaches to the correction of selectivity bias. *The Review of Economic Studies.* **49**(3), pp.355-372.

Line, N.D., Hanks, L. and Kim, W.G. 2016. Hedonic adaptation and satiation: Understanding switching behavior in the restaurant industry. *International Journal of Hospitality Management.* **52**, pp.143-153.

McAlister, L. 1982. A dynamic attribute satiation model of variety-seeking behavior. *Journal of consumer research.* **9**(2), pp.141-150.

McSweeney, F.K. and Swindell, S. 1999. General-process theories of motivation revisited: the role of habituation. *Psychological bulletin.* **125**(4), p437.

Murteira, J.M. and Ramalho, J.J. 2016. Regression analysis of multivariate fractional data. *Econometric Reviews.* **35**(4), pp.515-552.

Niyogi, R.K. 2015. *What to do, when to do it, how long to do it for: a normative microscopic approach to the labour leisure tradeoff*. thesis, UCL (University College London).

Palma, D., Enam, A., Hess, S., Calastri, C. and Crastes dit Sourd, R. 2021. Modelling multiple occurrences of activities during a day: an extension of the MDCEV model. *Transportmetrica B: Transport Dynamics.* **9**(1), pp.456-478.

Papke, L.E. and Wooldridge, J.M. 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics.* **11**(6), pp.619-632.

Park, J.-Y. and Jang, S.S. 2014. Revisit and satiation patterns: Are your restaurant customers satiated? *International Journal of Hospitality Management.* **38**, pp.20-29.

Pinjari, A.R. and Bhat, C. 2010. A multiple discrete–continuous nested extreme value (MDCNEV) model: Formulation and application to non-worker activity time-use and timing behavior on weekdays. *Transportation Research Part B: Methodological.* **44**(4), pp.562-583.

Pinjari, A.R., Bhat, C.R. and Hensher, D.A. 2009. Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B: Methodological.* **43**(7), pp.729-748.

Rajagopalan, B.S., Pinjari, A.R. and Bhat, C.R. 2009. Comprehensive model of worker nonwork-activity time use and timing behavior. *Transportation Research Record.* **2134**(1), pp.51-62.

Rasouli, S. and Timmermans, H. 2012. Uncertainty in travel demand forecasting models: literature review and research agenda. *Transportation letters*. **4**(1), pp.55-73.
Redden, J.P. 2008. Reducing satiation: The role of categorization level. *Journal of Consumer Research.* **34**(5), pp.624-634.
Saxena, S., Pinjari, A.R. and Bhat, C.R. 2021. Multiple discrete-continuous choice models with additively separable utility functions and linear utility on outside good: Model properties and characterization of demand functions.
Shabanpour, R., Golshani, N., Fasihozaman Langerudi, M. and Mohammadian, A. 2018. Planning in-home activities in the ADAPTS activity-based model: a joint model of activity type and duration. *International Journal of Urban Sciences.* **22**(2), pp.236-254.
Sobhani, A., Eluru, N. and Faghih-Imani, A. 2013. A latent segmentation based multiple discrete continuous extreme value model. *Transportation Research Part B: Methodological.* **58**, pp.154-169.
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**, pp.403-422.
Terza, J.V. and Wilson, P.W. 1990. Analyzing frequencies of several types of events: A mixed multinomial-Poisson approach. *The Review of Economics and Statistics.* pp.108-115.
Timmermans, H. 2005. *Progress in activity-based analysis.* Elsevier.
Varghese, V. and Jana, A. 2019. Multitasking during travel in Mumbai, India: Effect of satiation in heterogeneous urban settings. *Journal of Urban Planning and Development.* **145**(2), p04019002.
Victoriano, R., Paez, A. and Carrasco, J.-A. 2020. Time, space, money, and social interaction: Using machine learning to classify people's mobility strategies through four key dimensions. *Travel Behaviour and Society.* **20**, pp.1-11.
Wafa, Z., Bhat, C.R., Pendyala, R.M. and Garikapati, V.M. 2015. Latent-segmentation-based approach to investigating spatial transferability of activity-travel models. *Transportation Research Record.* **2493**(1), pp.136-144.
Ye, X. and Pendyala, R.M. 2005. A model of daily time use allocation using fractional logit methodology. In: *Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic TheoryUniversity of Maryland, College Park*.
You, S.I., Chow, J.Y. and Ritchie, S.G. 2016. Inverse vehicle routing for activity-based urban freight forecast modeling and city logistics. *Transportmetrica A: Transport Science.* **12**(7), pp.650-673.

Appendix C

Supplementary Table 5-1 Estimates from base FMNL and MDCEV model.

	Base FMNL			Base MDCEV	
	Estimate	Rob.t.rat.(0)		Estimate	Rob.t.rat.(0)
			Baseline parameter		
Basic needs	0	NA	Basic needs	0	NA
Household obligation	-1.885	-22.42	Household obligation	-2.556	-30.61
In-home work	-3.439	-16.78	In-home work	-4.409	-28.70
In-home education	-4.388	-13.42	In-home education	-5.528	-19.61
In-home leisure	-1.904	-25.45	In-home leisure	-2.480	-32.99
In-home social	-3.679	-30.91	In-home social	-4.161	-38.14
Work	-1.373	-17.51	Work	-2.901	-32.35
School	-3.635	-13.80	School	-4.991	-22.43
Personal issues	-3.362	-23.34	Personal issues	-3.726	-40.44
Shop	-3.268	-42.01	Shop	-2.933	-38.55
Leisure	-3.361	-26.75	Leisure	-4.204	-37.04
Social	-1.935	-30.87	Social	-2.502	-37
Accompany/Drop-off	-4.131	-29.33	Accompany/Drop-off	-3.866	-30.34
Travel	-1.988	-50.75	Travel	-0.779	-7.46
			Satiation parameter		
			Household obligation	1.786	15.46
			In-home work	2.463	6.98
			In-home education	3.646	6.18
			In-home leisure	1.658	17.52
			In-home social	1.494	13.66
			Work	6.444	13.86
			School	4.482	4.90
			Personal issues	1.141	12.14
			Shop	0.473	11.43
			Leisure	2.309	11.06
			Social	1.436	15.02
			Accompany/Drop-off	0.643	11.17
			Travel	0.257	10.57

LL (Start)	-3132.56	LL (Start)	-23504.71
LL (Final)	-2152.86	LL (Final)	-18144.53
AIC	4331.72	AIC	36343.07
BIC		BIC	36480.21

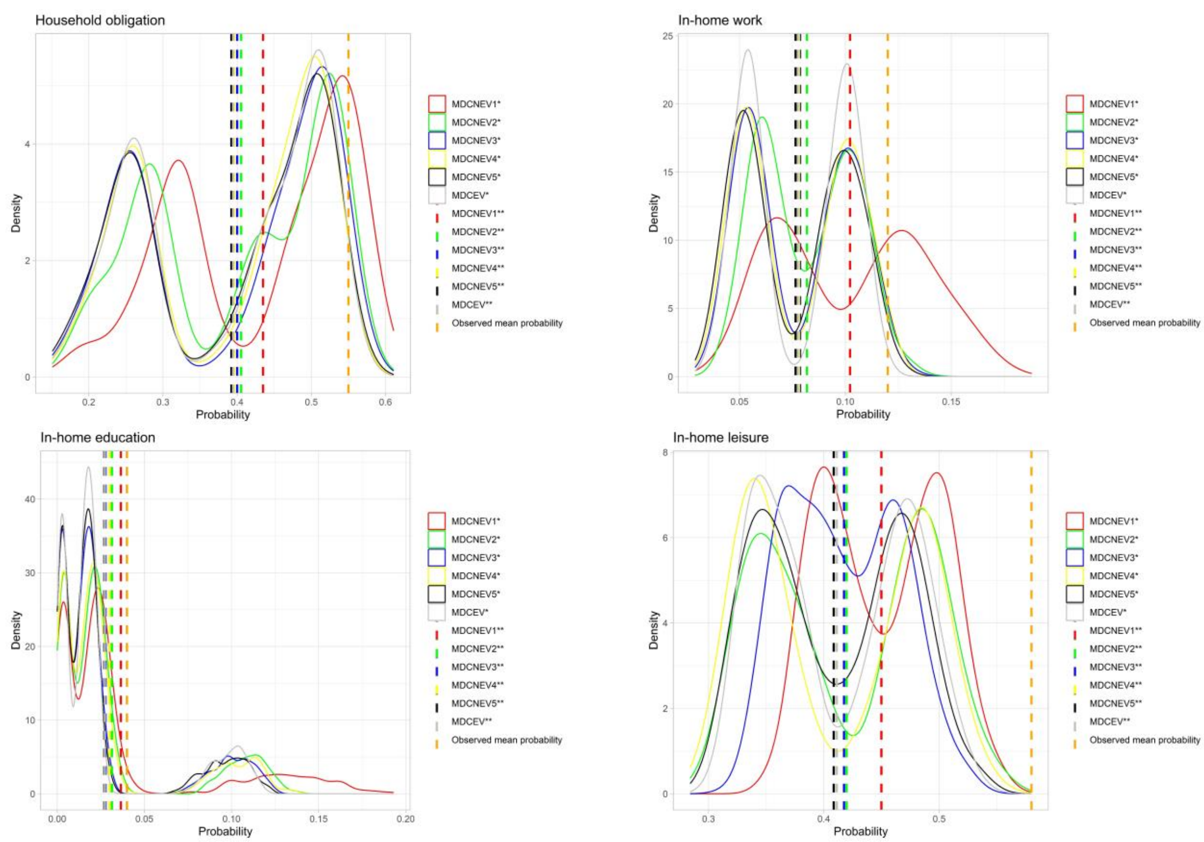
Supplementary Table 5-2 Prediction summary of base FMNL and MDCEV model.

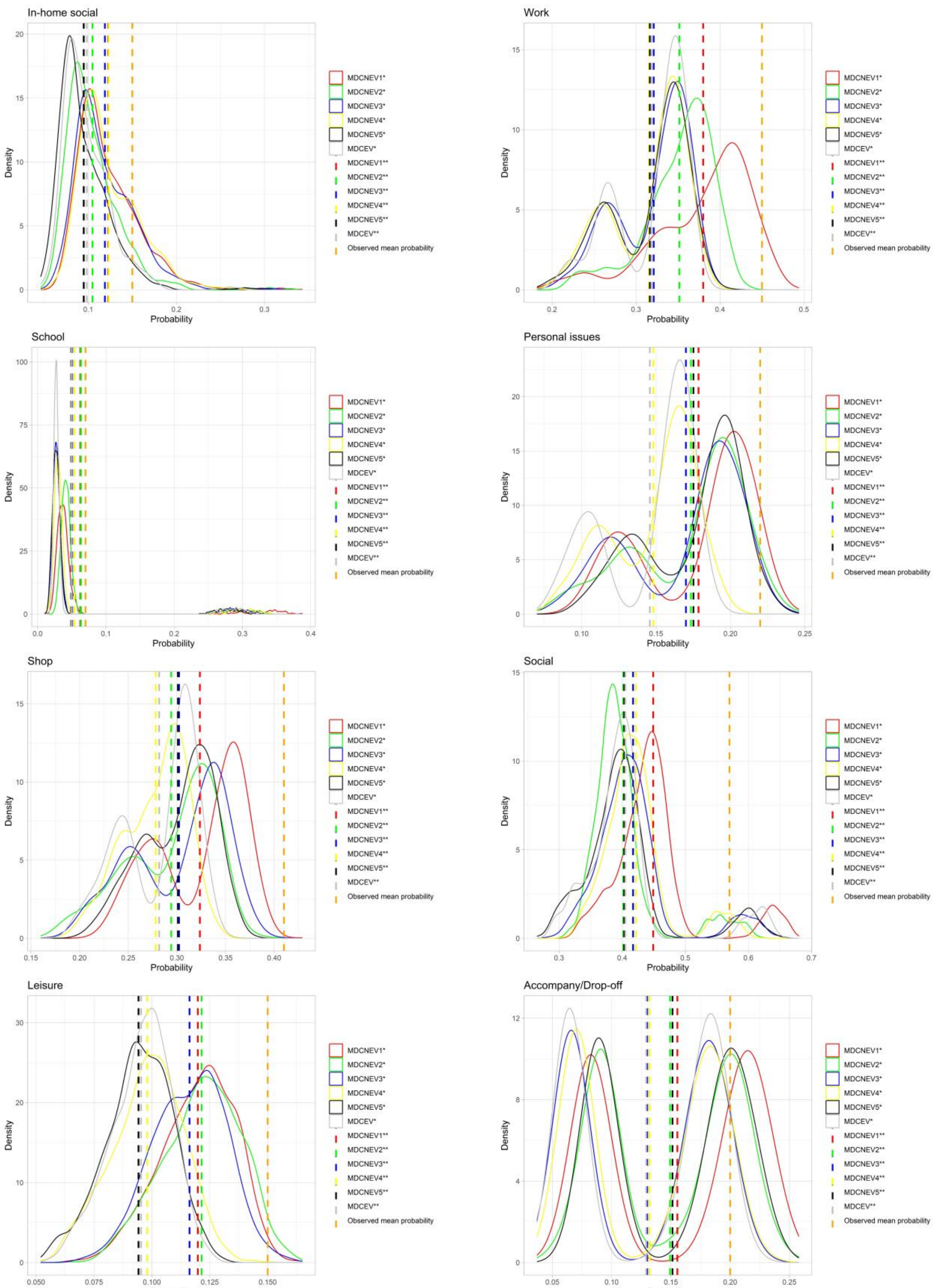
		FMNL	**MDCEV**	**MDCEV**	**FMNL**	**MDCEV**
	Observed time-use	**Predicted time-use**	**Predicted time-use**	**Predicted time-use (if participated)**	**RMSE**	**RMSE**
Basic needs	11.68	11.68	11.39	11.39	2.95	2.97
Household obligation	1.77	1.77	1.76	4.53	2.40	2.40
In-home work	0.37	0.37	0.35	4.45	1.34	1.34
In-home education	0.15	0.15	0.14	5.33	0.82	0.82
In-home leisure	1.74	1.74	1.80	4.41	2.27	2.27
In-home social	0.29	0.29	0.33	3.37	0.93	0.93
Work	2.96	2.96	2.48	7.78	3.76	3.79
School	0.31	0.31	0.27	5.91	1.37	1.37
Personal issues	0.40	0.40	0.43	2.94	1.17	1.17
Shop	0.44	0.44	0.51	1.82	0.91	0.91
Leisure	0.41	0.41	0.41	4.32	1.26	1.26
Social	1.69	1.69	1.62	4.04	2.46	2.46
Accompany/Drop-off	0.19	0.19	0.26	2.01	0.53	0.54
Travel	1.60	1.60	2.25	2.74	1.32	1.47
Overall					1.91	1.92

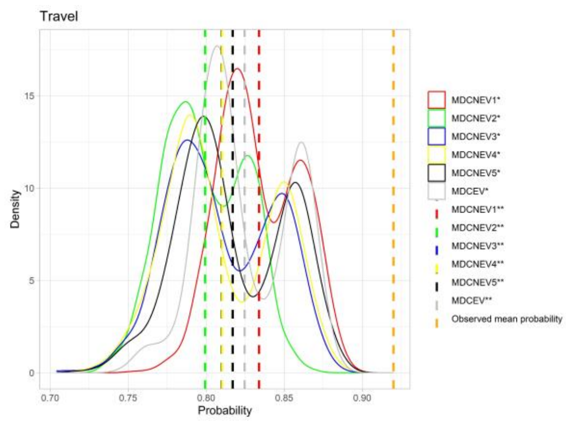
Supplementary Table 5-3 RMSE of the MDCEV and different MDCNEV models (if participated).

	MDCEV	MDCNEV 1	MDCNEV 2	MDCNEV 3	MDCNEV 4	MDCNEV 5
Basic needs	2.98	3.24	3.01	2.99	3.00	3.001
Household obligation	3.45	2.84	3.44	3.48	3.40	3.461
In-home work	4.28	3.22	4.46	4.41	4.44	4.346
In-home education	0.00	0.00	0.00	0.00	0.00	0.000
In-home leisure	3.45	3.60	3.48	3.41	3.43	3.468

In-home social	3.21	3.11	3.28	2.98	3.46	3.161
Work	6.07	5.55	6.25	6.09	6.08	6.058
School	5.59	4.58	5.36	5.60	5.74	5.617
Personal issues	2.79	2.72	2.37	2.60	2.81	2.996
Shop	1.64	1.62	**1.37**	1.51	1.62	1.746
Leisure	4.12	4.04	3.71	3.97	4.22	4.183
Social	3.38	3.47	3.23	3.33	3.54	3.373
Accompany/Drop-off	1.86	1.79	1.37	1.89	1.83	1.978
Travel	1.74	1.79	1.35	1.51	1.60	1.639




In-home social
Work
School
Personal issues
Shop
Social
Leisure
Accompany/Drop-off
Density
Probability
MDCNEV1*
MDCNEV2*
MDCNEV3*
MDCNEV4*
MDCNEV5*
MDCEV*
MDCNEV1**
MDCNEV2**
MDCNEV3**
MDCNEV4**
MDCNEV5**
MDCEV**
Observed mean probability



*Predicted discrete probability distribution
**Average mean probability
Supplementary Figure 5-1 Predicted discrete probability distribution of MDCEV and MDCNEV models.

Supplementary Table 5-4 Summary of t-test for different parameters of latent class MDCEV model.

	Class 1 vs. Class 2	**Class 3 vs. Class 1**	**Class 3 vs. Class 2**
	Diff-value (r t-ratio)	**Diff-value (r t-ratio)**	**Diff-value (r t-ratio)**
Baseline parameter			
Household obligation	0.863 (3.23)	-2.662 (-10.73)	-1.798 (-7.21)
In-home leisure	0.537 (2.61)	-1.429 (-4.92)	-0.892 (-2.90)
In-home social	0.704 (2.22)	-0.976 (-2.79)	-0.272 (-0.81)
Work	-2.492 (-7.29)	0.682 (1.23)	-1.810 (-4.79)
School			
Personal issues	0.744 (3.34)	-0.584 (-1.30)	0.159 (0.33)
Shop			
Leisure	0.281 (0.92)	0.780 (2.09)	1.062 (3.28)
Social	-0.379 (-1.92)	0.422 (2.12)	0.042 (0.18)
Satiation parameter			
Household obligation	-0.164 (-0.76)	0.370 (1.02)	0.206 (0.74)
In-home leisure	0.058 (0.27)	0.511 (1.47)	0.569 (1.84)
Work	-0.300 (-0.37)	4.225 (2.64)	3.926 (2.59)
School	3.077 (1.01)	-1.534 (-0.48)	1.543 (1.35)
Personal issues	0.731 (3.50)	-0.529 (-1.71)	0.203 (0.88)
Shop	0.262 (2.24)	-0.136 (-0.96)	0.126 (1.19)
Social	0.476 (1.87)	0.204 (0.52)	0.681 (2.09)
Accompany/Drop-off	0.220 (1.62)	-0.031 (-0.19)	0.189 (1.29)

Supplementary Table 5-5 Selected nesting structure and corresponding model performance (linear specification).

		Nest 1	Nest 2	Nest 3	LL(final)	AIC	BIC
1	Nest 1 (Obligation involved): Basic needs, household obligation, in-home work, in-home education, work, school Nest 2 (Less obligation involved): In-home leisure, in-home social, personal issues, shop, leisure, social, accompany/drop-off	0.383	1**		-2117.25	4300.51	
2	Nest 1 (In home activities): Household obligation, in-home work, in-home education, in-home leisure, in-home social Nest 2 (Out-of-home activities): Work, school, personal issues, shop, leisure, social, accompany/drop-off, travel	0.770**	1**		-2117.36	4300.71	
3	Nest 1 (Mandatory work): In-home work, Work Nest 2 (Discretionary activities): In-home leisure, in-home social, leisure, social, shop, personal issues, accompany/drop-off	0.553***	0.958***		-2117.17	4300.35	
4	Nest 1: In-home social, social, travel Nest 2: In-home work, work Nest 3: Household obligation, in-home education, in-home leisure, school, personal, shop, leisure, accompany/drop-off	0.366***	0.561**	0.968***	-2116.69	4301.38	
5	Nest 1: Personal, shop, accompany/drop-off, travel Nest 2: In-home social, social	0.327	(0.485)***		-2117.12	4300.24	

*** All the values are significant at 95% confidence level, ** All the values are significant at 90% confidence level, * All the values are significant at 80% confidence level, (…) at 95% confidence level values were significantly different from 1

Chapter 6 Developing an agent-based microsimulation for predicting the Bus Rapid Transit (BRT) demand in developing countries: a case study of Dhaka, Bangladesh

Abstract

Bus Rapid Transit (BRT) has been widely recognised as an affordable and effective mass transport system that can solve various mobility issues in countries that are unable to afford rail-based mass transit options. However, it is extremely challenging to predict the demand of the first BRT service in a city of a developing country with a weak public transport system using aggregate models, given the radical difference in the level-of-service between the BRT and the existing modes. Further, there can be substantial changes in the activity and travel patterns in a city after introduction of the BRT which simpler disaggregate level analysis tools are unable to predict. Agent-based simulation tools, which are the state-of-the-art tools for simulating complex travel behaviour, are hence more appropriate for predicting the network conditions after the introduction of a new BRT system. But the application of such simulation tools has been primarily limited to developed countries where the transport landscape and the travel behaviour are very different from the developing countries. To address this gap, this paper presents a demand forecasting model for BRT and integrates it in an activity-based micro-simulation tool in the context of Dhaka, the capital of Bangladesh and one of the fastest growing megacities in the world. The model was developed based on an existing multi-agent, activity-based, travel demand simulator (MATSim). The MATSim implementation in the context of Dhaka focused on two aspects: (1) implementing behaviour models in MATSim to reflect the mode choice in presence of the proposed BRT (2) integrating multiple data sources (including stated-preference data) for calibrating the mode choice and other components of MATSim to realistically mimic the travel behaviour in the city. Once calibrated, different access scenarios for BRT are simulated using MATSim and the sensitivity of the outputs to different modelling assumptions are tested. Results from the simulation showed that the marginal utility of travel time, travel cost and pricing structure of BRT significantly influenced BRT travel demands. Also, BRT demand was found to be the highest (25% of the total trips) in the scenario with multi-modal access/egress connections. While such direct model outputs presented in this paper will be useful for the planners to maximise the ridership of the proposed BRT, the calibrated simulator will be also useful for the evaluation of other innovative transport modes in the context of Dhaka in the future.

Keywords: Bus Rapid Transit, agent-based model, MATSim, public transport, Dhaka, Bangladesh, South Asia

6.1. Introduction

Rapid urbanisation and the associated increase in urban population in developing countries create significant pressure on transportation infrastructure (Makinde et al., 2018), causing various urban mobility challenges such as rising travel demand (Madlener and Sunak, 2011; Melo et al., 2012; Rahman, M.S.-U. et al., 2012), congestion (Han et al., 2019), safety issues (Cabrera-Arnau et al., 2020), increasing vehicle ownership (Cervero, 1996), and unreliable public transport (PT) services (Poku-Boansi and Marsden, 2018). To address such challenges, the government and policymakers of many countries have sought to prioritise PT in their transport policies (Mavi et al., 2018). In recent years, Bus Rapid Transit (BRT) has been widely recognised as an affordable and effective mass transport system that can solve various mobility issues in countries which are unable to afford rail-based mass transit options (Schalekamp and Behrens, 2010; Paget-Seekins, 2015; Venter et al., 2018; Joseph et al., 2021).

Although the BRT systems are in operation in many cities, issues related to planning and implementation of this mode of transport can lead to less successful outcomes (Deng and Nelson, 2013; Poku-Boansi and Marsden, 2018). A lack of knowledge of local settings (e.g., social-spatial system) could limit the success of BRT implementation in attracting people to this mode of transport (Joseph et al., 2021). Achieving a high modal shift to BRT is in fact a challenging process in developing countries due to the increasing income and affordability of private vehicles (e.g., cars and motorbikes) (Satiennam et al., 2016). Besides, predicting the demand for the first BRT service in a city with a weak public transport system is extremely challenging given the radical difference between the BRT and the existing modes. It requires collecting stated preference (SP) data which is rarely done in developing countries (Rastogi, 2000). Further, the spatial transferability performance of the models is typically not good in developing countries (Sanko, 2014; Bwambale et al., 2015). Therefore, to ensure substantial modal shift and achieve the environmental and mobility benefits resulting from BRT (Hidalgo and Graftieaux, 2008; Cervero, 2013), a city-specific empirical investigation of BRT operation is necessary (Joseph et al., 2021).

The existing studies conducted in the context of developing countries for planning BRT systems primarily adopted classical four-step models. However, the four-step models disregard the constellation of decisions of when, why, and where to travel (Hafezi et al., 2019; Västberg et al., 2020). Predicting the demands of a new transport system such as BRT using simple models is difficult, as the design and operation of BRT are complex, and require a wide range of control measures that are linked to the existing PT services and transportation infrastructure (e.g., rolling stock, right of way, pricing strategy, land use measures) (Paget-Seekins, 2015; McCormack et al., 2021). The agent-based travel demand model has emerged as a new generation of transport modelling and forecasting tools which provides an alternative to the traditional aggregate demand modelling. This modelling approach is flexible, comprehensive, and capable of modelling individual decision-making process. Alternatively, disaggregate models, like econometric models, hold the promises for forecasting demand for either existing or inaugural BRT (Chen and Naylor, 2011; Zgheib et al., 2020). However, to understand the changes in traffic conditions such as traffic flow, speed, congestion etc. relying solely on an econometric demand model proves inadequate (Zhang, L. and Levinson, 2004). On the other hand, agent-based microsimulation provides a valuable opportunity to comprehend how macro-level patterns like congestion arise from the individual decision of agents (Ausloos et al., 2015).

Agent based modelling framework (ABM) provides the flexibility to incorporate multiple attributes of agent and their environment in different layer/module format, and simulate the model to understand urban traffic flows (Grether et al., 2008; Manley et al., 2014; Małecki, 2018), activity behaviours (Arentze et al., 2010; Märki et al., 2014; Čertický et al., 2015; Gkiotsalitis and Stathopoulos, 2015; Shabanpour et al., 2017), changes in land use and effects on environment (Zhang, S. and Zhao, 2018), performance assessment of service (Gao et al., 2016; Ji et al., 2018; Levine et al., 2018), accessibility of location (Huang, 2019), location decision of housing (Ding et al., 2018), joint travel mode and departure time choice (McDonnell and Zellner, 2011; Zou et al., 2016; Jing et al., 2018), joint route choice and departure time choice (Li et al., 2018) and many more. While there have been several attempts to adopt agent-based models for understanding modern complex transport-related issues, those were primarily limited to larger metropolitan planning organisations (MPOs) in developed countries. The application of an agent-based model in fast-growing cities of developing countries is very challenging due to required input data (e.g., travel information, census data and infrastructure related data), and computational challenges (Kagho et al., 2020).

To address issues regarding travel demand prediction in developing countries, this paper presents a demand forecasting model for BRT and integrates it into an agent-based micro-simulation tool. The model was established in the context of Dhaka, the capital of Bangladesh and one of the fastest-growing megacities in South Asia and the world. An existing multi-agent, activity-based, travel demand simulator (MATSim) has been utilised in this regard. The MATSim implementation in the context of Dhaka focuses on two aspects: (1) implementing behaviour models in MATSim to reflect the mode choice in presence of the proposed BRT and (2) integrating multiple data sources (including stated-preference data) for calibrating the mode choice and other components of MATSim to realistically mimic the travel behaviour in the city. Finally, the impacts of the two different accessibility scenarios (with and without multimodal feeder service) and two different pricing structure (monthly flat fare and distance-based fare system) were quantified using the calibrated and validated simulator.

The remaining of the article is organised as follows: first, we provide a short description of the RP and SP data used in this study, followed by a description of the modelling framework. The results from the simulation are then presented with a critical discussion of this study and future research direction.

6.2. Data

6.2.1. Study area

This study focused on the Dhaka Metropolitan Region (hereinafter RAJUK area) (Figure 6-1). Dhaka, the capital of Bangladesh, is home to more than 15 million people. The population of the city is likely to increase to 26.3 million by 2035, predominantly due to rural-urban migration (DTCA, 2016). To meet the growing mobility demands in this city, the rate of car ownership, as well as the growth of motorised vehicles, have been increasing at an alarming rate. According to the Bangladesh Road Transport Authority (BRTA), 918,233 private motorised vehicles (cars, jeeps, microbuses and motorcycles) were newly registered in Dhaka between 2011-21, while the corresponding number for buses and minibuses was 23,887 (BRTA 2022). Furthermore, there is a particularly inefficient use of road space due to the low occupancy rate of private vehicles. The escalating growth in motorised vehicles, low road capacity, and a lack of traffic management resulted in significant traffic congestion, causing a loss of 3.2 million business hours per day in Dhaka (Siddique et al., 2017). The city's administration does not, however, maintain or organise them efficiently (Sajib, 2021). In Dhaka, public transport service can be characterized by no regular schedule and fixed vehicle stops, overcrowded services, frequent service suspension, limited traffic service monitoring (Quddus et al., 2019; Satu and Chiu, 2019; Rahman, F., 2022). Lack of adequate public transportation facilities to fulfil passengers' mobility demand is now a serious concern in this mega city. The World Bank reported that during the morning and evening peak hours, the average vehicle speed is approximately 8.75km/h in the road network (WB, 2015). The mean travel time in peak hours is almost three times higher than the travel time in off-peak hours.

To reduce the level of congestion, the government invested heavily in new infrastructure (e.g., flyovers, BRT, and MRT services) in the last two decades. However, most of the structural interventions have failed to minimise traffic congestion, primarily due to a lack of system-level analyses and the absence of robust transport models (Habib, 2002; Enam, 2010). The Strategic Transport Plan of 2005 (DTCB, 2005) recommended implementing three BRT lines. Among those, line 3 is currently under construction and is expected to be in full operation by the end of 2025 (Supplementary Figure 6-5). The BRT lines will be operating in parallel to the existing PT services (Figure *6-2*). The mode choice demand for BRT was estimated only by calibrating PT demand using conventional four-stage modelling while ignoring its potential impact on the existing transport services (e.g., auto-rickshaw, motorcycle, human-hauler), infrastructure, and land use (ADB, 2022).

Considering Dhaka city as a case study area and the under-construction BRT (Line 3 in the Strategic Transport Plan 2005) as a new transport option, this study focused on developing a simulation environment to assess mode choice behaviour and the impact of BRT in the existing setting. The proposed model aimed to predict the mode choice at an individual scale, over different periods subject to spatial and temporal constraints.

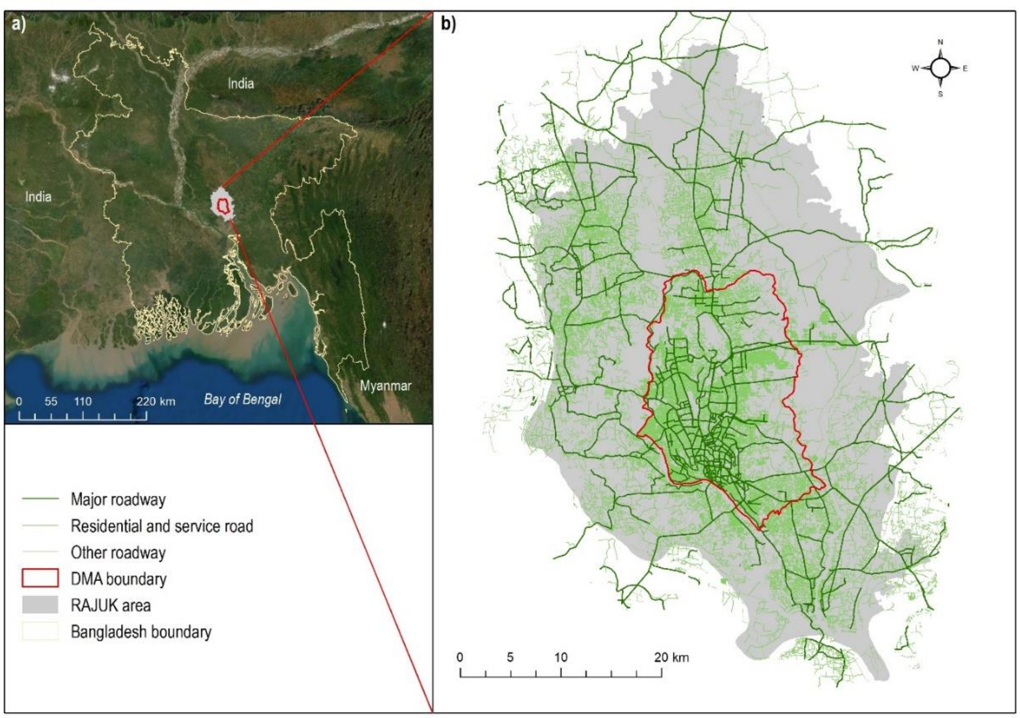


Figure 6-1 Location of the study area with the existing road network.

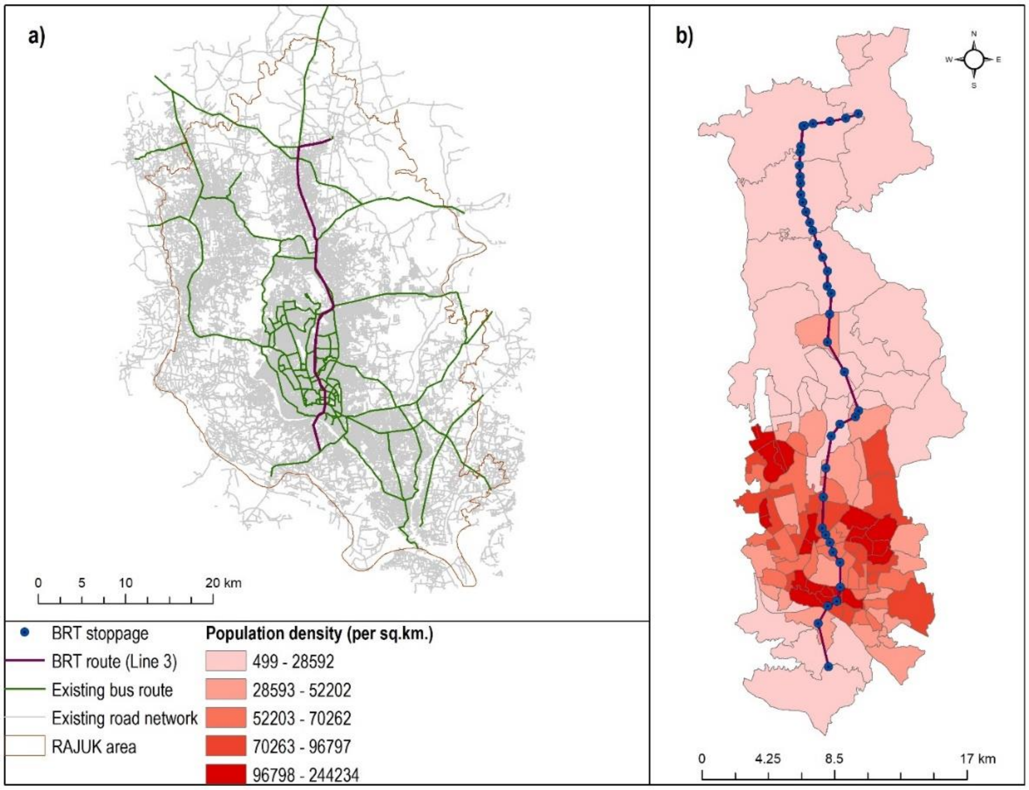


Figure 6-2 Population distribution and Transit network.

This study utilised both Revealed Preference (RP) and Stated Preference (SP) data for generating the MATSim inputs and development of the mode choice models. The datasets are described below.

6.2.2. Revealed Preference (RP) Data

This study utilised RP data collected from the Dhaka Transport Coordination Authority (DTCA), which commissioned a consulting firm TYPSA (www.typsa.com) to develop the database by carrying out a travel diary survey across the RAJUK area as part of an ongoing Dhaka Subway Project. The survey was conducted from Monday to Saturday between 28th February 2019 to 4th May 2019. The survey form included two main sections. The first part was related to general household characteristics (e.g., age, gender, education, occupation, income, car ownership). The second section included trip-related information (e.g., departure time, travel mode, travel time, trip purpose) of each member of all sample households, who made trips on the previous working day. Information on commuting, education, leisure, personal and other purpose travel was collected. A total of 35,000 households were surveyed in the RAJUK Area, which constitutes approximately 1% of the total households in Dhaka. The details related to survey, data and summary can be found in the “Feasibility study and preliminary design for construction of Dhaka subway” report (TYPSA, 2019). The data summary of RP data is shown in Table 6-1.

6.2.3. Stated Preference (SP) Data

Since BRT is an ongoing project, solely the RP data is unable to determine the sensitivity toward attributes (e.g., travel time, travel cost) of this new transport mode. Many studies combined RP and SP data either to analyse the impact or to evaluate the consumer preference of multi-attributed product or services (Ben-Akiva and Morikawa, 1990; Brownstone et al., 2000; Bhat and Castelar, 2002; Rashedi et al., 2017). This is due to the fact that context effects are limited in a realistically designed SP choice experiments since SP data are not based on actual market behaviour (Ben-Akiva et al., 1994). Therefore, we also used SP data of 1,016 individuals, collected by Enam (2010) using a roadside questionnaire survey (conducted between December 2009 and January 2010). The questionnaire comprised of three sections: 1) questions related to current travel patterns (e.g., origin, destination, mode choice, trip purpose, duration, distance travelled), 2) choice of alternative in hypothetical scenario (the choice task included respondents' current mode of transport and two improved mass transport alternatives: BRT and MRT), and 3) socio-demographic characteristics (e.g., age, gender, occupation, education, household income, car ownership). In this study, our primary focus was on the hypothetical scenario presented with different level-of-service attributes (e.g., travel time, cost) using pictorial representations. Respondents were asked if they would continue the current mode or shift to BRT or MRT in each scenario. Further details on survey design, data collection and summary findings can be found in Enam (2010). In our study, the SP choice data collected by Enam (2010) and RP data collected by TYPSA (2019) was used to develop a coupled RP-SP model to understand the preference for BRT and MRT and to improve the accuracy of the parameter estimates. Given the scope of the current work being limited to BRT, only the components of the joint model related to the choice among the existing modes (walk, bike, rickshaw, bus, human hauler, motorcycle, auto-rickshaw) and BRT have been used for the simulation of the future scenarios.

Table 6-1 RP and SP data summary.

Socio-demographic variables	*SP (%)*	*RP (%)*
Age		
<18	1.73	12.87
18-25	42.44	17
25-40	42.44	38.48
40-60	12.95	26.68
>=60	0.44	4.97
Gender		
Male	76.41	73.71
Female	23.59	26.29
Household size		
<=5	78.55	88.44
>5	21.45	11.56
Car Ownership		
No car	71.01	93.82
1+ car	28.99	6.18
Occupation		

Student	32.94	24.55
Office-employee	46.33	36.64
Self-employed personnel	12.2	24.31
Housewife	5.72	9.47
Retired	0.65	1.66
Unemployed	1.3	0.75
Other	0.86	2.61
Education		
Below SSC	4.22	36.25
SSC	4.88	15.34
HSC	18.83	19.41
BSC	45.24	13.47
MSC or above	24.89	13.42
Other	1.94	2.11
Income (in BDT)		
<10,000	6.04	1.89
10,000-20,000	15.4	15.16
20,000-30,000	25	24.97
30,000-40,000	16.38	23.47
>=60,000	37.18	34.51

6.3. MATSim framework

MATSim is an open-source multiagent traffic simulation platform that consists of several modules, which have been used as a standalone system to implement a large-scale agent-based transport model in several studies (Raney et al., 2003; Balmer et al., 2006; Balmer et al., 2009; Vosooghi et al., 2017). This queue-based network simulation platform is the most widely used agent-based simulator in the field of transportation (Balmer et al., 2006; Bouman et al., 2012).

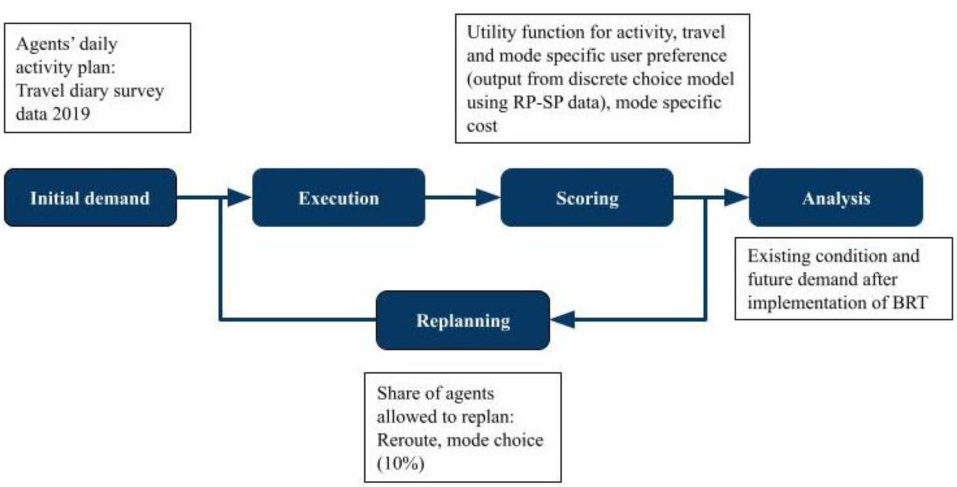


Figure 6-3 MATSim simulation framework used in this study.

Figure 6-3 represents an adaptation of the framework of the MATSim simulator for this particular study. The simulation sets off with each agent's initial day plan encoded from the daily activity chains of the population of the study area. The virtual world represents the transport infrastructure (road network, service facilities) and land use. Each agent within the simulator can perform their daily activity plan in the corresponding virtual world. With iteration-by-iteration, the initial demand is optimized in mobsim[25] based on its associated score (equivalent to 'utility function' used in a random utility based econometric framework). This iteration process is repeated until the average score reaches the stable condition. For scoring agents' plans, several parameters can be used for the (dis)utility associated with travelling, waiting, performing an activity etc. After each iteration, agents' plans are scored to eliminate bad plans (i.e., plans with lower utility) so that only viable day plans could evolve to further iteration steps in the simulation. Further, during the iteration agents are allowed to modify their plan, known as replanning[26]. At this stage, a certain share of agents can change their departure time, route, mode, and location of some activities.

To predict the BRT demand, the overall simulation process was organized into four major steps — simulation inputs preparation, scoring parameter estimation for mode assignment, scoring of agent's activity plan, and running the simulation scenario.

6.3.1. Simulation inputs

MATSIM, in its simplest form, requires three types of inputs: 1) activity plan of the population of the area (or a representative fraction of the population), 2) transport network with the description of the road and 3) configuration (which dictates the specific demand modelling process).

6.3.1.1. Activity profile

Activity demand can be generated either by translating the total population census or by generating a synthetic population using a sample (Axhausen et al., 2016). Since the latest census data available for Dhaka was collected in 2011, we opted for the option to generate activity plans using microdata from a representative sample of the population. The household-level trip diary data from the TYPSA survey (see 2.2 for details) was used in this regard. However, the trip diary only included the detailed geo-location of each participant's household, and the locations of the different activities were only available at the TAZ level. The locations of various activities were randomly assigned within the TAZ boundary using the Geographic Information System (GIS) in a manner that they match with the user stated travel time (Bekhor et al., 2010). After determining location information, an activity profile was generated based on information extracted from travel diary survey data. Each activity plan included information about activity location (x-y coordinate), end time of first activity, 'leg' mode and maximum duration allocated for that activity.

[25] Mobsim is the mobility simulation module in MATSim. Two internal mobsims — QSim and JDEQSim, are available in the MATSim default library. External mobility simulations can also be plugged into MATSim (Axhausen et al., 2016).

26 At the replanning stage, each agent can select a strategy for plan selection (best score/logit model/random etc.) and/or an innovation strategy (mode choice, route choice, departure time choice), where a certain plan of an agent is updated. Each strategy is given a weight determining the probability, by which the course of action represented by that strategy is taken. For multiple strategies, weights are normalized. More details about each strategy can be found in Axhausen et al. (2016).

6.3.1.2. Infrastructure

To represent a virtual urban transport landscape of Dhaka, this study utilised a road network and available transport services. The road network for the study area was obtained from the Open Street Map (OSM) service[27]. In MATSim, the available modes are car, public transport (PT), bike, and walk. Other modes such as auto-rickshaw and motorbike, which exist in Dhaka city, are not in-built alternatives in MATSim. They were modelled utilising special vehicular specifications within the existing framework in MATSim. These vehicular specifications are presented in Table 6-2. Since the data represents 1% of the total households in Dhaka City, the supply side adjustment was done in MATSim by considering flow capacity and storage capacity factor of 0.015 (Mehlstäubler, 2019). The outflow capacity of a link — the number of travellers leaving the respective link per time step — is defined by the 'flow capacity'. The number of cars fitting onto a network link per time step (Axhausen et al., 2016) is defined by the 'storage capacity'.

Table 6-2 Vehicular specification for MATSim implementation.

Mode of transport	Length (meter)	Width (meter)	Maximum velocity (m/s)	PCE
Auto-rickshaw (CNG)	2.7	1	16	1
Motorbike	2.2	1	22.22	0.5
BRT	18	2.5	25	3

Source: Kadiyali (2010), the PCE value of BRT was assumed to be the same as a Tram of similar dimensions

In the latest OSM map, BRT line 3 was not part of the existing road network. Since BRT is an ongoing project, in this study, we developed an artificial network of BRT line 3 and added it to the existing OSM road network. To artificially replicate the BRT operation with the existing network, the following criteria were considered for BRT scheduling and vehicle definition:

- BRT line 3 stations were introduced at the proposed locations.
- BRT route was coded to have a dedicated right-of-way (grade-separated).
- The service frequency of BRT was defined from the standard available in the BRT feasibility report (three minutes intervals from 8:00 to 10:00 and 16:00 to 17:00 and 10 minutes intervals for the remaining time of the day) (ADB, 2022).
- BRT was assumed to be available for all residents of the city.

6.3.1.3. Configuration

The default configuration settings of the MATSim simulation were used as the starting point in this study. However, in the default module, access/egress mode for PT is only limited to walking. However, in Dhaka, rickshaws are the most widely used access/egress modes. Besides, a considerable number of people use walking and cycling for short-distance trips. Hence, in the multimodal access/egress scenario for this study, walking, rickshaw, and bike were considered. Additional search radius parameters were specified for rickshaws and bike: if no PT stops were found within the initial radius (4km), the search radius expanded until a stop was found (up to the maximum search radius of 6km).

[27] All reproducible codes (e.g., generation of existing road network from Open Street Map, BRT network, schedules etc.) can be found in the following link: khzannat/dhaka-matsim-example-project (github.com). Activity plan file is not shared due to data privacy.

6.3.2. Mode assignment

To estimate the scoring parameters for mode assignment, mode choice models were developed following the random utility framework. The developed mode choice models predicted the choice between existing modes (e.g., car, bus, rickshaw, human hauler, auto-rickshaw, motorbike, walk, bike) and BRT. The random utility theory suggests that individual decision is followed by rationality and complete information (McFadden, 1973). Agents choose each alternative transport mode with the highest utility, where the utility of an alternative *i* to a person *n* has the following form:

$$u_n(i) = u(x_{in}, s_n) \tag{1}$$

where x_{in} is the vector of the attribute of alternative i for individual n and s_n is the vector of characteristics of the person n.

McFadden (1973) proposed that this utility has the linear-in-parameters separable form presented below:

$$u(x_{in}, s_n) = V(x_{in}, s_n) + \varepsilon_{in} \tag{2}$$

where V is the observed component of utility. The unobserved variable ε_{in} represents the random error term. (McFadden, 1973). The choice probabilities for each alternative i in MNL can be expressed as follows (for detail see (Train, 2009)):

$$P_n(i) = \frac{e^{V_{in}}}{\sum_{j=1}^{j} e^{V_{jn}}} \tag{3}$$

In evaluating the existing scenario, none of the available modes showed correlations in the error term. Therefore, we developed a multinomial logit model to estimate the mode-specific constant and analyse travel time sensitivity, where the error term ε_{in} is independently and identically distributed (IID).

Furthermore, as BRT information was not available in the RP data and using a model-based only on SP data may be prone to hypothetical bias, we developed a joint RP-SP model. As proposed by Ben-Akiva and Morikawa (1990), the difference between the error terms in RP and SP can be modelled as a function of the variances of each type of error and can be presented as follows:

$$\sigma_{RP}^2 = \mu^2 \sigma_{SP}^2 \tag{4}$$

where μ is the scale coefficient.

After adopting the formulation for RP and SP data the utility equation can be written as follows:

$$u^{RP}({x_{in}}^{RP}, s_n) = V^{RP}({x_{in}}^{RP}, s_n) + {\varepsilon_{in}}^{RP} \tag{5}$$

$$\mu * u^{SP}({x_{in}}^{SP}, s_n) = \mu * (V^{SP}({x_{in}}^{SP}, s_n) + {\varepsilon_{in}}^{SP}) \tag{6}$$

Probability of choosing alternative *i* among the available alternative *j* in the RP data:

$$P_n^{RP}(i) = \frac{e^{V_{in}^{RP}}}{\sum_{j=1}^{J} e^{V_{jn}^{RP}}} \tag{7}$$

Probability of choosing alternative *i* among the available alternative *j* in the SP data*:*

$$P_n^{SP}(i) = \frac{e^{\mu V_{in}^{SP}}}{\sum_{j=1}^{J} e^{\mu V_{jn}^{SP}}} \tag{8}$$

Joint log-likelihood function:

$$LL(\beta) = \left(\sum_{n=1}^{N}\sum_{i} y_{ni}^{RP}\, ln\,(P_{ni}^{RP})\right) * \left(\sum_{n=1}^{N}\sum_{i} y_{ni}^{SP}\, ln\,(P_{ni}^{SP})\right) \tag{9}$$

The coefficients of the joint model were estimated using the maximum likelihood technique using the package Apollo in R programming language (Hess and Palma, 2019). In the combined model, we used SP and RP specific alternative specific constants. Car travel time for RP data was measured using Google Maps API. Using the same tool, time specific congestion factors were calculated for different origin destination pairs. Obtained congestion factors and mode specific travel time at free flow speed were used to estimate the travel time of other alternative motorised modes (Bwambale et al., 2019). Travel cost for RP data was measured using a distance cost multiplier. We estimated the unknown utility function parameters using the combined model where the common parameter was the travel time and travel cost in our model. Estimates of the common parameters as well as any RP- and SP-specific model parameters were available through joint estimation of the two models. A "scale" parameter was used to equalise the scale of the coefficients of the two models because the variances of the random components of the RP and SP utility functions were likely to differ. The joint model was used for simulating the existing and future mode choice scenarios. The estimated model parameters were used as the scoring parameters (e.g., generic marginal utility of travel time, marginal utility of money) for the simulation. Since using a bike or walking involves kinetic energy rather than a direct monetary cost, we calculated how sensitive walking and cycling are to distance. In order to investigate the effects of the inaugural BRT on different market shares, we also estimated the time sensitivity of different socio-demographic groups (e.g., male vs. female, employed vs. unemployed, working aged group vs. other age group, having a car at the household's vs no car). The model summary used for the simulation is presented in Table 6-3.

Table 6-3 Discrete choice model.

Parameters	MNL 1			MNL 2		
	Estimate	Robust *t*-stat	Sig.	Estimate	Robust *t*-stat	Sig.
Alternative Specific Constants (ASCs)						
RP specific parameters						
Walking	2.094	35.990	***	2.050	35.608	***
Bike	-0.794	-9.685	***	-0.836	-10.182	***
Rickshaw	1.677	40.000	***	1.648	39.280	***
Bus	2.170	45.400	***	2.128	44.451	***
Human hauler	-0.102	-1.923	**	-0.144	-2.721	***
Motorcycle	0.153	2.967	***	0.113	2.188	***
Auto-rickshaw	0.261	5.872	***	0.212	4.717	***
Car	0	-	-	0	-	-
SP specific parameters						
Rickshaw	-0.439	-1.736	**	-0.762	-2.963	***
Bus	-1.822	-8.468	***	-2.161	-9.929	***
Auto-rickshaw	0.081	0.274		-0.311	-1.091	***

Car	0	-	-	0	-	
BRT	-0.728	-3.786	***	-0.781	-3.891	***
MRT	-1.064	-5.252	***	-1.106	-5.308	***
Level of Service Attributes						
Generic travel time (per hour)	-0.802	-20.753	***			
Generic travel cost (BDT)	-0.002	-7.093	***	-0.003	-8.137	***
Distance sensitivity for walking (m)	-0.125	-15.619	***	-0.128	-16.229	***
Distance sensitivity for cycling (m)	-0.076	-8.834	***	-0.077	-8.941	***
Market specific travel time						
Travel time (base)				-0.309	-5.712	***
Employed				-0.103	-1.729	***
Having car at household				-1.358	-10.670	***
Male				-0.288	-4.709	***
Age between 24 to 40 years old				-0.389	-7.703	***
Scale variables						
mu_RP	1			1		
mu_SP	1.694	3.748	***	1.274	4.355	***
LL(0, whole model)	-209245.3			-209245.3		
LL(final, whole model)	-157093.1			-156657.1		
LL(final RP only model)	-156173.8			-155734.9		
LL(final SP only model)	-919.3			-922.17		
Rho-square (0)	0.2492			0.2513		
Adj.Rho-square (0)	0.2492			0.2512		
AIC	314220			313356.2		
BIC	314382			313556		
Estimated parameters	17			21		

6.3.3. Scoring of agent's activity plan

In this study, one virtual day was iteratively simulated for a 1% sample of Dhaka city. During the iteration process, a predefined number of agents were allowed to change some of their daily decisions to search for a plan with a higher utility. All agents tried to adapt their plans in such a way that their utility is improved by keeping the track of each activity chain. The model was run until the population reached an equilibrium condition (Balmer et al., 2006; Bouman et al., 2012). The plan of each agent at the equilibrium condition was a plausible approximation of the real-world behaviour of an individual.

The optimization process described above was based on the evaluation of the plans using a specific scoring function. The MATSim scoring function used in this research was formulated by Charypar and Nagel (2005), loosely based on the Vickrey model for road congestion (Vickrey, 1969). The utility of a plan U_{plan} is estimated as the sum of all activity utilities $U_{act,q}$ plus the sum of all travel (dis)utilities $U_{trav,mode(q)}$ as presented below:

$$U_{plan} = \sum_{q=0}^{N-1} U_{act,q} + \sum_{q=0}^{N-1} U_{trav,mode(q)} \tag{13}$$

where N is the number of activities and trip q is the trip that follows activity q.
Following the scoring function, the utility of an activity q is calculated as follows:

$$U_{act,q} = U_{dur,q} + U_{wait,q} + U_{late,q} + U_{early,q} + U_{s.dur,q} \tag{14}$$

where $U_{dur,q}$, is the utility of performing activity q, $U_{wait,q}$ denotes waiting time spent in front of the closed activity location, $U_{late,q}$ specifies the late arrival penalty, $U_{early,q}$ defines the penalty for not staying long enough and $U_{s.dur,q}$ is the penalty for a 'too short' activity. This study hypothesised that marginal utility of activity duration will be decreasing logarithmically (the detail will be found in Axhausen et al. (2016)).
The disutility associated with the travel for a leg q is given as:

$$S_{trav,q} = ASC_{mode\,(q)} + \beta_{trav,mode(q)} * t_{trav,q} + (\beta_{distnace,mode(q)} + \beta_{cost,mode(q)} * C_{mode\,(q)})\, d_{trav\,(q)} \tag{15}$$

where $ASC_{mode\,(q)}$ is the mode specific constant, $\beta_{trav,mode(q)}$ is marginal utility of time spent traveling by mode, $t_{trav,q}$ is the travel time between activity locations q and $q + 1$. $\beta_{distnace,mode(q)}$ is the marginal utility of distance included in a direct manner for walking and bike use (as these mode requires physical effort). $ASC_{mode\,(q)}$, $\beta_{trav,mode(q)}$, $\beta_{cost,mode(q)}$, and $\beta_{distnace,mode(q)}$ were derived from MNL model (Table 6-3).

6.3.4. Other simulation settings

The available modes to the agents at the base scenario were car, bus, rickshaw, human hauler, auto-rickshaw, motorbike, walk, and bike (Supplementary Figure 6-4 shows photographs of different transport modes available in the study area). For the future BRT scenario, BRT was added as an additional alternative to the existing modes. The 'real' travel times of car, motorcycle, auto-rickshaw, and BRT were obtained through the traffic simulation component. Due to the exclusion of the narrow roads in the available network files (which are applicable for walking, cycling, and rickshaw), the non-motorised modes were simulated using 'adjusted beeline distances' with the following specifications through the 'teleportation module' within MATSim. Due to the absence of reliable information about the bus and human-hauler routes, they were also simulated using the 'teleportation'[28] feature, but with network-derived travel distances.
For the simulation, two different fare scenarios have been tested: a monthly flat fare and a distance-based cost scenario (cost/km). For the monthly flat fare, the ticket price was assumed to be 900 BDT per month (as it is estimated that about 50% of the commuters spend less than 900 BDT in a month) (ADB, 2022). For the distance-based cost, the distance travelled between activity locations q and $q + 1$ were derived from the activity profiles and the travel cost per kilometre 1.52 BDT/km was derived from the feasibility report of BRT. All the additional simulation settings are specified in Table 6-4.

[28] Teleportation is a method of moving vehicles from origin to destination, at a predefined speed, without considering interactions in the network.

Table 6-4 Simulation settings.

Mode	Speed	Distance	Cost
Bus	16 km/hr	Network-derived	1.52 BDT/km
Rickshaw	10km/hr	2* bee-line distance	10 BDT/km
Human-hauler	13km/hr	Network-derived	2 BDT/km
Auto-rickshaw	Network derived	Network-derived	8 BDT/km
Car	Network derived	Network-derived	6 BDT/km
Motorcycle	Network derived	Network-derived	6 BDT/km
Walk	3km/hr	1.4* bee-line distance	
Bike	8km/hr	1.3* bee-line distance	

6.3.5. Simulation scenario

Three different scenarios were simulated for comparison — 1) Base scenario without BRT, 2) Scenario with BRT without multi-modal access, and 3) Scenario with BRT with multimodal access. A summary of various scenarios used to simulate traffic is presented in Table 6-5.

Table 6-5 Summary of different scenarios.

	Scenarios	Input	Scoring
1	Base scenario (without BRT)	- Existing road network of Dhaka city - One day activity profile - Configuration	MNL 1 model
2	Future scenario with BRT	- Existing road network of Dhaka city with BRT line 3 and stoppage - One day activity profile - Configuration without multimodal accessibility	MNL 1 model MNL 2 model (for different market share)
3	Future scenario with BRT	- Existing road network of Dhaka city with BRT line 3 and stoppage - One day activity profile - Configuration with multimodal accessibility	MNL 1 model MNL 2 model (for different market share)

6.4. Results

6.4.1. Scenario 1: Base scenario

The base scenario of the proposed mode choice model reflects the existing conditions (pre-BRT scenario) of modal share, trip purpose, and departure time choice. In this scenario, agents were allowed to change the route to obtain the shortest path. To start with a more stable base model, we also allowed our agents to change one single trip mode (randomly picked) till agents reached their equilibrium. We compared the simulated modal share and departure time choice with the observed data obtained from the travel diary survey conducted in 2019. For the validation, the simulated modal share was also compared with the modal share of the passenger trips collected using an inner cordon line survey, in 2014 by the JICA study team. The results indicated that public transport and non-motorised transport, constituted the highest share of trips in Dhaka City. As Table 6-6 exhibits, the simulated proportion of modal share has a good agreement with the observed and JICA data. However, a difference between the observed and simulated proportion of the modal share of the agents was discernible for existing NMT and Auto-rickshaw, which is approximately 19% and 10%, respectively. We accepted these differences due to the validation of these differences in the external sources. Also, it should be worth mentioning the fact that rickshaws and bicycles are not legally permitted along the major roads in Dhaka. We attempted to restrict those vehicles along the major road while routing without network simulation. However, due to a lack of law enforcement, people can use those prohibited vehicles along the major road while violating the law.

Table 6-6 Observed vs simulated modal share.

	Observed passenger modal share (%)	**Simulated passenger modal share (%)**	**JICA survey (2014) passenger modal share (%)**
Bus and Human-hauler	49.2	53.0	68.4
NMT (walking/rickshaw/bike)	34.8	16.0	13.4
Car	3.7	11.0	8.2
Auto-rickshaw	7.0	17.0	8.4
Motorbike	5.3	3.0	1.6

The base model also showed the agents' choice of departure time and trip purpose in existing situations. During a weekday, for both home-based outbound and return trips, the number of mandatory trips (i.e., work and education) (more than 80% trips) was the highest in Dhaka city. However, the proportion of trips for personal reasons, leisure activities, shopping, and other purposes was substantially low during the weekday. In comparison to home-based travels, the proportion of non-home-based trips was extremely small, accounting for less than 5% of all observed trips. In the case of departure time choice, most of the agents started their trip before 9:00 for work and educational purposes. However, departure times for other activities peaked at 10:00. (Figure 6-4).

a) Mandatory trips (work and education)

b) Non-discretionary trip (personal, social, leisure and other activities)

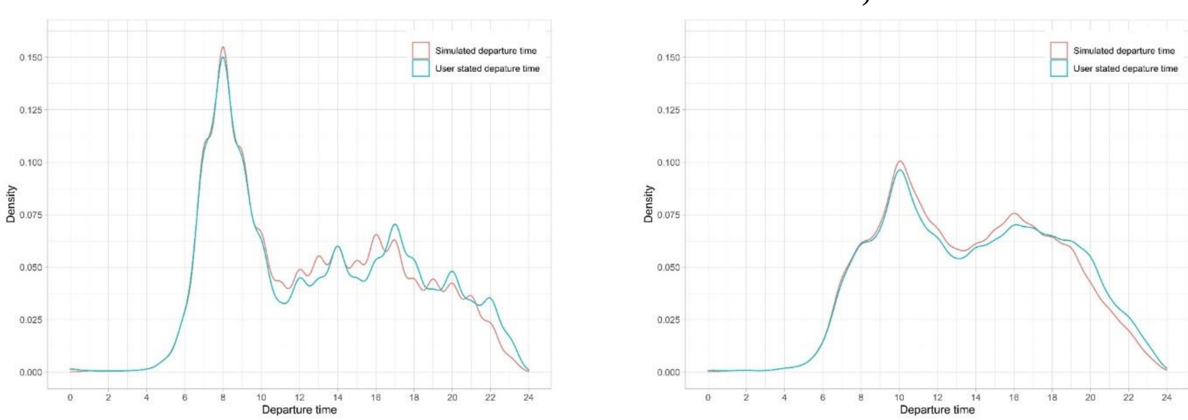


Figure 6-4 Observed vs Simulated departure time density at the base scenario for different trip purposes.

The distribution of simulated travel time was further compared against the user stated travel time. Figure 6-5 shows that in both the observation and simulation, the majority of agents' travel duration ranges from

20 to 40 minutes. Such a validation process yielded a very good agreement between the observed and modelled travel time distributions (Figure 6-5).

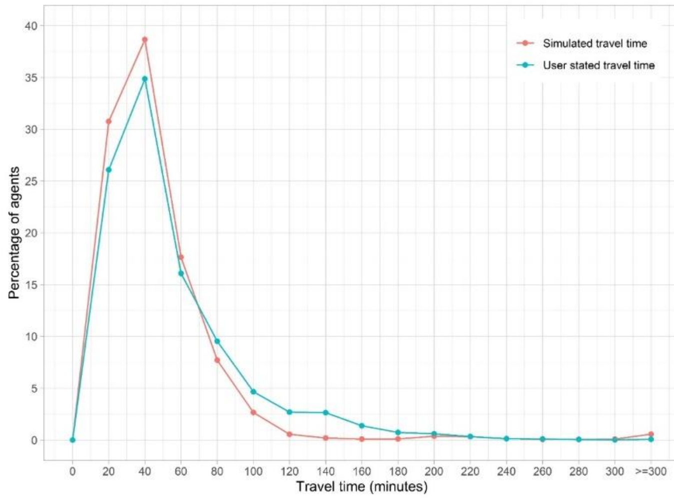


Figure 6-5 Comparison of travel time in simulation and user stated travel time.

6.4.2. Future mode choice scenarios

This study analysed sensitivity to travel time, travel cost, and multimodal access-egress modes in post-BRT implementation scenarios. Each of these three variables was included in the model stepwise.

6.4.2.1. Scenario 2: Future Scenario with BRT and without multi-modal accessibility

In scenario 2, we intended to see how travel time and travel cost will influence agents' mode choice behaviour after the implementation of BRT. Here, to avail of BRT, agents could only use walking as an access/egress mode. We tested the influence of travel time independently, as well as in combination with travel cost as a function of distance. For scoring in the simulation, the mode choice model provided the marginal utility of travel time and travel cost and distance disutility for a mode which required physical effort of the agent (e.g., walking and bike).

The optimisation results of travel time showed that if all else being equal, agents chose the mode of transport that yielded the minimum travel time. In this case, the motorbike resulted in the highest modal share (approximately 65%) at the equilibrium point (scenario 2A, Figure 6-6 (a)). Since, results from the discrete choice model highlighted the significance of both time and cost sensitivity in predicting agents' mode choice preference, in this study, we also investigated the trade-off between travel time and travel cost. Results showed that agents chose the mode of transport that required the least travel cost and travel time. The highest modal share was obtained in the bus at the equilibrium point (Figure 6-7 (a) and Supplementary Figure *6-1*). The implementation of BRT with walk as an access mode would attract 1.2% of the total users from other modes, particularly buses, rickshaws, and motorcycles (scenario 2B). A decrease in the link flow of network simulated motorised modes (car, motorcycle, and auto-rickshaw) was noticeable along the road parallel to the BRT network (Dhaka-Mymensingh highway) at various times of the day, however, there was also an increase in link flow at various times in different links (Supplementary Figure 6-3).

From the simulation of this scenario, it can be found that a significant proportion of trips (64.2%) could still take place on buses, followed by auto-rickshaws (15.2%) (Figure 6-7 (b)). The average travel distance of BRT users would be 10 km (sd ± 6.9km). We have also examined the impact of a BRT service's flat monthly fare on demand (scenario 2C). The demand for BRT would marginally increase to 1.4% with this fare structure. The simulation result, however, indicated a considerable increase in the average journey distance of potential BRT users (11km, sd ± 7.8km). On the contrary, in the distance-based cost simulation, agents would avail each BRT stoppage on an average by walking approximately 650m (sd ±500m), whereas, for the flat fare cost simulation agents avail each stoppage by travelling 550m (sd ± 420m).

When the simulation considered travel cost as a function of distance, the proportions of home-based and non-home-based trips of the potential BRT users were 91.5% and 8.5%, respectively (scenario 2B). Among the home-based trips, 65.3% were work-related. In this case, non-home-based work trips were 77.9% (Supplementary Table 6-1). Similarly, at the simulation of monthly flat fare of BRT service, about 67.5% of trips were home-based work trips, while 80.0% were non-home-based work trips (scenario 2C). In both cost scenarios, the BRT users would prefer to depart either during the morning (7:00 – 9:00) or afternoon (16:00 – 18:00) (Figure 6-7 (c)). Considering departure time, multiple peaks were observed in a working day (Figure 6-7 (c)).

In this study, we also simulated the potential demand for BRT while taking into account the varying temporal sensitivity of various market shares. According to the simulation results, the potential demand for BRT varies between 0.6% to 0.7% for different market share. The majority of potential BRT users would be those who were younger than 25 or older than 40 (40% of 0.6% were aged between 25 to 40 years old), employed (73.5% of 0.7% BRT users were employed), male (80.3% of 0.7% BRT users were male), and did not have a private car in the household (91.7% of 0.6% BRT users did not have a car) (Supplementary Table 6-3).

a) b)

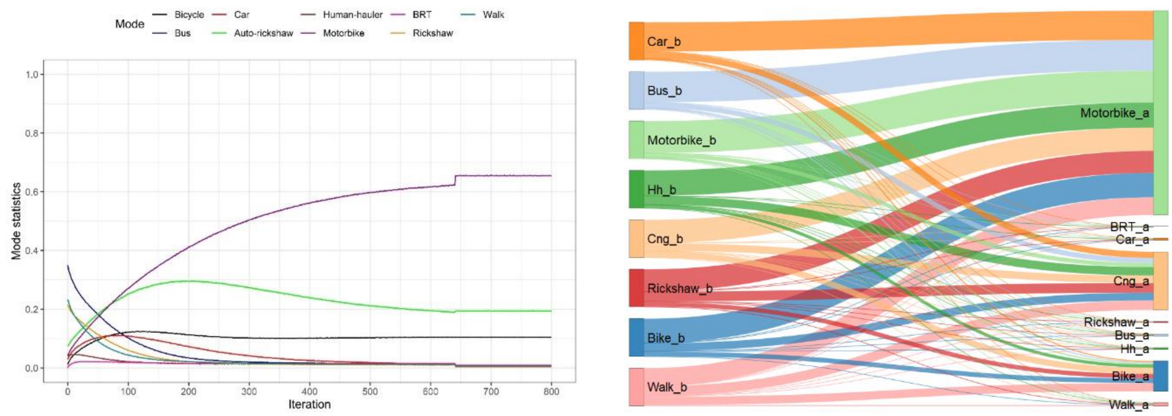


Figure 6-6 a) Optimization of travel time across various modes; b) Modal shift based on travel time preference.

a) b)

c)

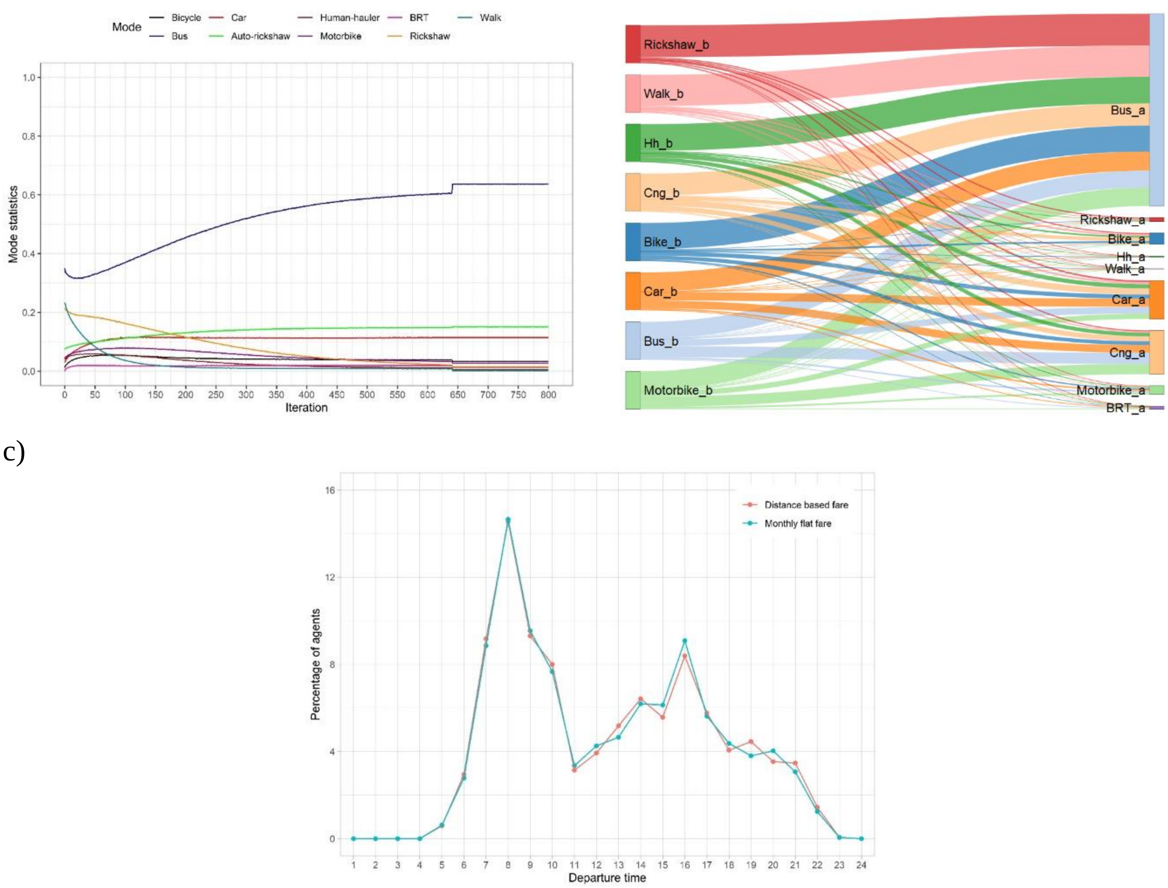


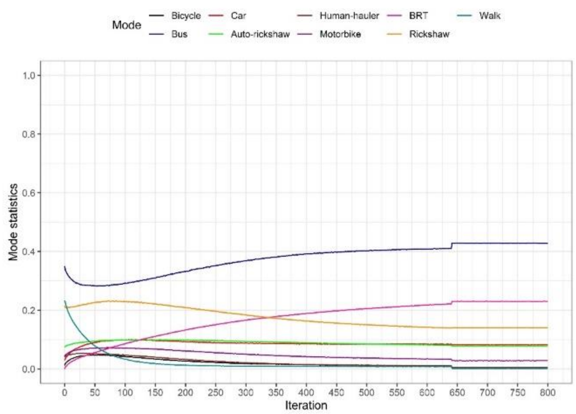
Figure 6-7 a) Optimisation of travel time and travel cost across various modes; b) Modal shift based on travel cost (function of distance) preference; c) Departure time choice of potential BRT users under different pricing structure of BRT.

6.4.2.2. Scenario 3: Future Scenario with BRT and multi-modal accessibility

The third scenario considered the travel time and travel cost of users, as well as the presence of multimodal access-egress modes. The distribution of modal share at equilibrium points (Figure 6-8 and Supplementary Figure 6-2) under this scenario indicated that users would still use bus as their preferred mode of transport. The implementation of BRT with multimodal accessibility could attract approximately 23% of the total users from other modes (e.g., bus, rickshaws, and motorbikes) in both fare systems. This scenario also noted a significant reduction in the link flow along the existing road parallel to BRT line (Supplementary Figure 6-3). In this scenario (3A), about 60.5% of the total potential BRT users would be the existing bus users (Figure 6-8 (b)), followed by rickshaw (11.8%) and motorbikes (7.3%). Approximately similar result obtained from the simulation of monthly flat fare system with multi-modal transport accessibility (Supplementary Figure 6-2).

In both fare system with multi-modal transport accessibility, the proportion of home-based trips was approximately 93% whereas the highest proportion of BRT trips were work (63%) and education (21%) related (Supplementary Table 6-1). Among the non-home-based trips, majority of the trips were work related trips (73%) which was followed by personal activities (15%), and shopping (5%). It is noteworthy that the BRT users could travel approximately 12 km (sd ± 6km), on average, in different fare system. In both fare systems, most of them would prefer to depart between 7:00 to 10:00 (Supplementary Table 6-2). Furthermore, the proportions of the walk, bike, and rickshaw usage, as access modes, for any of the trip legs of BRT users were found to be 1.3%, 7.6%, and 91.1%, respectively. In different multi-modal simulation scenario, agents were availing BRT service by travelling approximately 1.7km (sd ± 1.3km) using different non-motorised transport allowed as access/egress mode. While considering the different travel time sensitivity of different market share in the multi-modal accessibility scenario, BRT service would attract approximately 24 to 25% of the total users from other modes. Like the scenario 2, among these potential BRT user majority of them were employed, younger than 25 or older than 40, male, and did not have a private car in the household (Supplementary Table 6-3).

a)



b)

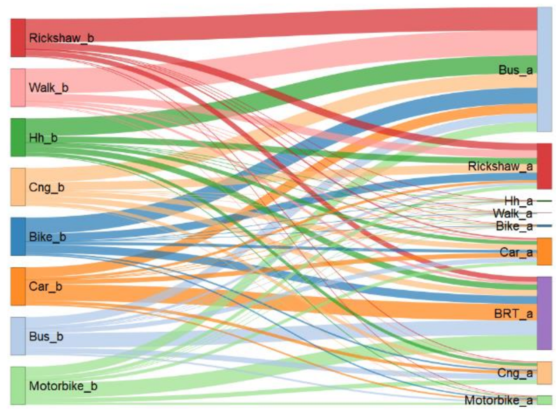


Figure 6-8 a) Distribution of modal share at multi-modal accessibility scenario; b) Modal shift based on the presence of multimodal access-egress modes.

6.5. Discussion of results and policy implications

Transportation demand prediction is essential to evaluate the investment in future transport infrastructure. This is particularly important for a mega project like BRT since investment in such a project has both success and failure reports. Agent-based microsimulation approach has received wider attention recently, to forecast travel behaviour (Makinde et al., 2018). While the application of this approach in evaluating transport service/infrastructure is common in developed countries (Moreno et al., 2018; Manser et al., 2020), there is a lack of attempt to adopt such a method in developing countries, limiting the effectiveness of new transport infrastructure and service (Yagi and Mohammadian, 2010). This research gap led to this study where we developed an agent-based model to predict future travel demands of an ongoing Bus Rapid Transit (BRT) project in Dhaka city of Bangladesh. In this study, a base scenario was developed to artificially represent the existing travel pattern in Dhaka city, combining an estimated discrete choice model with a MATSim interface that merged both the demand and supply sides. Results from the base model showed that public transport services (bus, human-hauler) and non-motorised transport (walking and rickshaw) constituted the highest share of trips. This simulated result complies with the observation and survey result from the most recently available cordon line survey data collected by the JICA study team (DTCA, 2016).

The calibrated agent-based-microsimulation tool was then applied to simulate and compare the future travel demand of BRT in three scenarios: (1) existing scenario, (2) future scenario with BRT and without multimodal accessibility, and (3) future scenario with BRT and with multimodal accessibility. The results indicated that if agents are only sensitive to travel time, they will choose the quickest mode for their daily trips. After the implementation of BRT line 3, the share of trips by motorcycle increased by 65% when agents were only time-sensitive (scenario 2A). In Dhaka city, the number of yearly registered motorcycles increased from 34,707 in 2011 to 99,810 in 2021 (BRTA, 2022), primarily due to its flexible route choice, door-to-door access, and relatively lower travel time than other alternatives (Wadud, 2020). Besides, the expected total journey time by BRT was greater than the motorcycles, despite the higher average travel speed of BRT as a journey by BRT involved access and egress time (Mavi et al., 2018; Zgheib et al., 2020; Shi et al., 2021), making it less attractive than motorcycles which can provide seamless door-to-door trip. Therefore, in order to shift people to BRT, the transport policy should include measures to make motorcycles less attractive — by increasing the import duties of motorcycles or by restricting them on the routes competing with BRT. Given that motorcycles currently have significant detrimental effects on air quality (Chiou et al., 2009) and road safety (Wadud, 2020), such policies can contribute to improving transport sustainability.

The simulation results further showed that after the implementation of BRT, a substantial proportion of passenger trips would still be based on buses when agents are both travel time and travel cost sensitive (scenario 2B and scenario 2C). While BRT intends to optimise both travel time and travel cost, this study only considered BRT line 3, which connects the north and south parts of Dhaka. Hence, dwellers in the rest of the city area would still be dependent on the existing PT services to fulfil their mobility demands. Therefore, to increase the PT ridership, revitalising the existing PT and enhancing inter-modal connectivity with BRT are imperative along with the construction of new BRT lines (Duarte and Rojas, 2012). The importance of intermodal connectivity is also supported by the result from scenario 3. The outputs of that scenario indicated that the presence of rickshaws or bikes as access/egress modes would substantially increase BRT service areas (from 0.7% to 25%), increasing the number of long-distance trips (Figure 6-9 shows the potential trip distribution of BRT users). These findings may have two important implications: 1) There is a strong likelihood that BRT would be appealing to city dwellers as an efficient and effective mode of public transportation to meet the current passenger travel demand if BRT has a promising schedule, connectivity, and service frequency. The efficient operation may reduce the demand for other motorised mode of transport (Supplementary Figure 6-3). It may increase travel demand along the access/egress connection street to accommodate first and last mile trips, potentially shifting congestion from major streets to the connecting street. BRT demand could decrease to 0.7% if link street is unable to meet the anticipated access/egress demand. The significance of connection and level of service features in meeting passenger transport demand by BRT was also highlighted in the empirical research by Joseph et al. (2021). Additionally, it was noticed from the simulation of scenario 2 that there was a possibility of an increase in link flow along the parallel road of the BRT network at various times of the day (Supplementary Figure 6-3). This is because switching to BRT would relieve traffic on the parallel road, which might then make it the quickest path for passengers travelling between other origin-destinations. A further increase in link flow may result in jam density instead of relieving congestion.

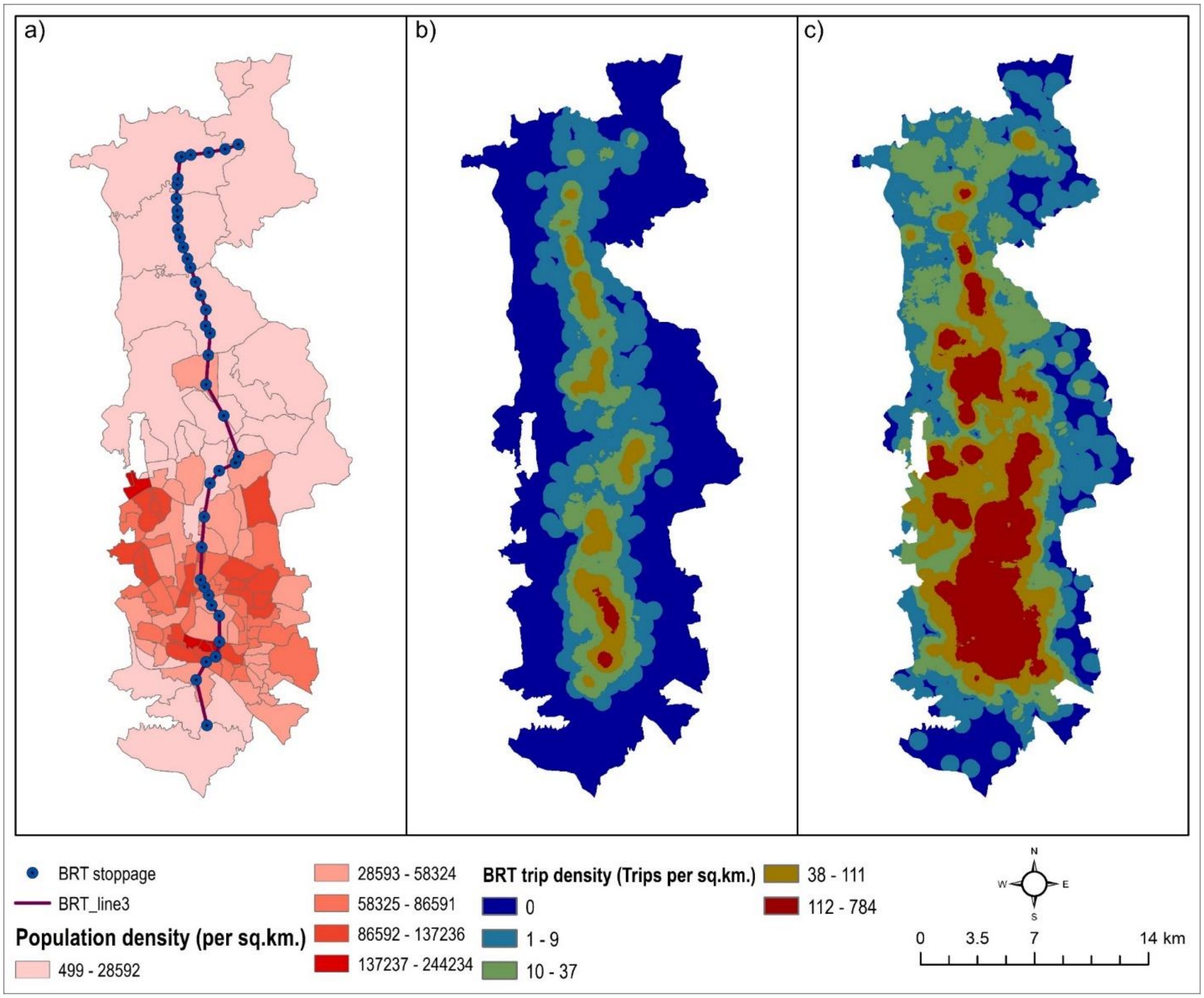


Figure 6-9 Potential BRT users' trip distribution pattern (a) BRT network, b) trip density without multi-modal connectivity, c) trip density with multi-modal connectivity).

This study also found the sensitivity of the fare structure on the BRT travel demands. Though consideration of walking as an access mode with monthly flat fare of BRT had induced a slight increase in BRT demand (1.2% to 1.4%), however, significance of different fare structure was discernible while evaluating the average journey distance of BRT users. This is because, monthly flat fares induced more long-distance trips (average journey distance increased from 10km to 11km with sd ± 7.8km). Similarly, the result showed that multi-modal accessibility further increased the average journey distance (scenario 3). Therefore, the availability of different fare schemes would be effective in increasing BRT ridership (particularly long-distance traveller) and reducing congestion along the busy corridor. It may be noted that Yagi and Mohammadian (2008) also highlighted the significance of different fare scheme in increasing BRT ridership in Jakarta, Indonesia, using an opinion survey.

In terms of trip purposes, home-based work trips during the weekdays will constitute the majority of the BRT trips. This indicates the need for taking special policy interventions to make BRT similarly attractive for discretionary trips. Potential measures may include an introduction of group tickets, integrated fare systems, and periodical (weekly and monthly passes) and multi-trip tickets could encourage people to use BRT (Currie and Delbosc, 2013; Currie and Delbosc, 2014). Furthermore, results from this study highlighted that demand for BRT would vary across different socio-demographic groups due to their different sensitivity to different level of service attributes. For example: due to different time sensitivity the demand for BRT would be dominated by the employment status, gender, and car ownership at the household level. Therefore, BRT ridership enhancement policies should include customised services (e.g., commuters' trip tickets, workplace incentives) or targeted marketing approaches (e.g., gender-sensitive planning), to better align with the preferences and sensitivities of distinct socio-demographic groups.

6.6. Conclusion

This study predicted the BRT demand of Dhaka city, using an agent-based micro-simulation approach. A mode choice model was developed using both RP and SP data which was implemented in MATSim. The developed models worked reasonably well among all the dimensions (travel time, travel cost and access/egress mode) considered in this study. In terms of practical application, the model developed in this study may help to understand the activity patterns and travel behaviour of the traveller after the initiation of a new BRT service. The Strategic Transport Plan (STP) (2005–2025) and (DTCA, 2015) predicted that the expected modal share of BRT in Dhaka city would be 3% for the proposed three BRT lines (Ahmed et al., 2018). But the predicted BRT demand in this study ranged between 0.7% and 25 % while taking into account the marginal utility of travel time, various combinations of fare structure, and multi-modal access/egress connection. In the STP, only public transport simulation was conducted, considering walking as an access mode. Such demand prediction model ignored the potential shifts from other modes (e.g., rickshaw, motorbike, human hauler), which have been considered in this study. The findings of the current study are hence expected to produce more reliable results. The key ridership enhancement policies inferred from the findings of this study are listed below:

1. Successful implementation of large mega-projects like BRT should consider the competitive advantages of other transportation modes, motorcycles in particular. Policy measures to make motorcycles less attractive, by increasing the import duties of motorcycles or by restricting them on the routes competing with BRT for instance, are hence crucial for shifting people to BRT.
2. Integration with current transportation services and intermodal connectivity is a prerequisite for increasing the BRT system. In particular, taking policy measures to promote the use of rickshaws and bikes as access/egress modes will substantially increase BRT service areas and help to increase the number of long-distance trips. It is crucial to take measures to ensure the link streets to the BRT line are capable of meeting the access/egress demand of the feeder modes (rickshaws and bikes).
3. Given the heterogeneous sensitivity to the level-of-service variables, policies aimed at increasing BRT ridership should incorporate tailored services (like workplace incentives and commuter trip tickets) and focused planning and marketing strategies (like gender-sensitive planning and promotion). These will ensure that BRT is able to better cater to the sensitivities of various socio-demographic groups and improve ridership.

Future research directions

While the direct model outputs presented in this paper will be useful for the planners to maximise the ridership of the proposed BRT, the calibrated simulator will be useful for evaluating strategies related to BRT ridership and innovative transport modes in the context of Dhaka. Some of the potential future research directions are outlined here:

The agent-based multimodal simulation only tested a limited number of level-of-service attributes (e.g., fare values already proposed by the consultants, no consideration for special fare types for to attracting discretionary travel, integrated ticket system with regular PT, etc.). Testing wider ranges of values and different combinations of level of service attributes can help to identify further strategies for improving BRT ridership. Also, other choice behaviour such as departure time choice and destination choice behaviour would be worthwhile to explore because ignoring the full range of behavioural changes may lead to over/underestimation of the potential benefits of the mega project.

Additionally, once the BRT is operational, it will potentially change the current interaction pattern of different types of travellers (such as new transit users, conventional transit users, and users of other modes of transportation) and service providers, law enforcement authorities, policymakers, etc. involved in the existing system (Palacios et al., 2020; Joseph et al., 2021). Due to the chaining and feedback effect, such interactions between agents and the transport system can have an impact on other system components (such as land use, economic conditions, etc.) while agents will be making decisions about their daily activities and mobility (e.g., what activity, when, where, what mode of transport) (Venter et al., 2018; Zgheib et al., 2020). In parallel, it will change the landscape of the city by enhancing connectivity, sprawling, gentrification, densification, land use change, and many more. Such a feedback loop might result in nonlinear causality within the adapted or evolved system bringing radical change in how agents act and how travellers behave (Ettema et al., 2014). Therefore, in future research, our current model can be elevated to a more dynamic model by incorporating the emerging interaction of individual agents with other agents and the environment.

It is expected that after the implementation of policy intervention (e.g., BRT, MRT, expressway, etc.), the current or preferred choice will be also affected by their habits, awareness, evolving attitude, culture, social norms and values (Zmud and Sener, 2017; Shafi et al., 2022; Wee and Kroesen, 2022). If a new BRT line is introduced and users find the service to be convenient and easy to use, this may transform their perception of public transportation and encourage them to utilise the service more regularly (Ramos et al., 2019). Similarly, adaptation to this new service may change their current social belief and perspective (Forward, 2019). However, it is extremely challenging to capture detailed behavioural nuances like exploration, habit formation, inertia, etc. using stated preference data. Extending the current model to capture these behavioural nuances using revealed data collected after opening of the BRT will be an interesting direction of future research. Such a model will also be useful for ex-post evaluation of the BRT system.

However, the way MATSim was adapted for use in the multi-modal context of a developing country is likely to be useful for other countries interested in transitioning to transport policy evaluations using agent-based microsimulation. It may be noted that the developed model is available as open-source software for ease of application in other cities and countries.

6.7. References

ADB. 2022. *Preparing The Greater Dhaka Sustainable Urban Transport Corridor. Final report* Dhaka, Bangladesh Asian Development Bank (ADB).

Ahmed, S., Nahiduzzaman, K.M. and Hasan, M.M.U. 2018. Dhaka, Bangladesh: Unpacking challenges and reflecting on unjust transitions. *Cities.* **77**, pp.142-157.

Arentze, T.A., Ettema, D. and Timmermans, H.J. 2010. Incorporating time and income constraints in dynamic agent-based models of activity generation and time use: approach and illustration. *Transportation research part C: emerging technologies.* **18**(1), pp.71-83.

Ausloos, M., Dawid, H. and Merlone, U. 2015. Spatial interactions in agent-based modeling. *Complexity and geographical economics: Topics and tools.* pp.353-377.

Axhausen, K.W., Horni, A. and Nagel, K. 2016. *The multi-agent transport simulation MATSim.* Ubiquity Press.

Balmer, M., Axhausen, K.W. and Nagel, K. 2006. Agent-based demand-modeling framework for large-scale microsimulations. *Transportation Research Record.* **1985**(1), pp.125-134.

Balmer, M., Rieser, M., Meister, K., Charypar, D., Lefebvre, N. and Nagel, K. 2009. MATSim-T: Architecture and simulation times. *Multi-agent systems for traffic and transportation engineering.* IGI Global, pp.57-78.

Bekhor, S., Dobler, C. and Axhausen, K.W. 2010. Integration of activity-based with agent-based models: an example from the tel aviv model and MATSim. *Arbeitsberichte Verkehrs-und Raumplanung.* **628**.

Ben-Akiva, M., Bradley, M., Morikawa, T., Benjamin, J., Novak, T., Oppewal, H. and Rao, V. 1994. Combining revealed and stated preferences data. *Marketing Letters.* **5**(4), pp.335-349.

Ben-Akiva, M. and Morikawa, T. 1990. Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General.* **24**(6), pp.485-495.

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), pp.593-616.

Bouman, P., Lovric, M., Li, T., Van Der Hurk, E., Kroon, L. and Vervest, P. 2012. Recognizing demand patterns from smart card data for agent-based micro-simulation of public transport.

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

BRTA. 2022. *Number of registered vehicles in whole Bangladesh.* [Online]. [Accessed May]. Available from: http://www.brta.gov.bd/site/page/74b2a5c3-60cb-4d3c-a699-e2988fed84b2/Number-of-registered-Vehicles-in-Whole-BD

Bwambale, A., Choudhury, C.F. and Hess, S. 2019. Modelling departure time choice using mobile phone data. *Transportation research part A: policy and practice.* **130**, pp.424-439.

Bwambale, A., Choudhury, C.F. and Sanko, N. 2015. *Modelling car trip generation in the developing world: The tale of two cities.*

Cabrera-Arnau, C., Prieto Curiel, R. and Bishop, S. 2020. Uncovering the behaviour of road accidents in urban areas. *Royal Society open science.* **7**(4), p191739.

Čertický, M., Drchal, J., Cuchý, M. and Jakob, M. 2015. Fully agent-based simulation model of multimodal mobility in European cities. In: *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*: IEEE, pp.229-236.

Cervero, R. 1996. Mixed land-uses and commuting: Evidence from the American Housing Survey. *Transportation Research Part A: Policy and Practice.* **30**(5), pp.361-377.

Cervero, R. 2013. *Bus rapid transit (BRT): An efficient and competitive mode of public transport.* Working paper.

Charypar, D. and Nagel, K. 2005. Generating complete all-day activity plans with genetic algorithms. *Transportation.* **32**(4), pp.369-397.

Chen, C.-H.P. and Naylor, G.A. 2011. Development of a Mode Choice Model for Bus Rapid Transit in Santa Clara County, California. *Journal of Public Transportation.* **14**(4), pp.1-22.

Chiou, Y.-C., Wen, C.-H., Tsai, S.-H. and Wang, W.-Y. 2009. Integrated modeling of car/motorcycle ownership, type and usage for estimating energy consumption and emissions. *Transportation Research Part A: Policy and Practice.* **43**(7), pp.665-684.

Currie, G. and Delbosc, A. 2013. Exploring comparative ridership drivers of bus rapid transit and light rail transit routes. *Journal of Public Transportation.* **16**(2), p3.

Currie, G. and Delbosc, A. 2014. Assessing bus rapid transit system performance in Australasia. *Research in Transportation Economics.* **48**, pp.142-151.

Deng, T. and Nelson, J.D. 2013. Bus Rapid Transit implementation in Beijing: An evaluation of performance and impacts. *Research in Transportation Economics.* **39**(1), pp.108-113.

Ding, H., Yang, M., Wang, W. and Xu, C. 2018. Simulating and analyzing the effect on travel behavior of residential relocation and corresponding traffic demand management strategies. *Ksce Journal of Civil Engineering.* **22**(2), pp.837-849.

DTCA. 2015. *Revised strategic transport plan (RSTP), Dhaka, Bangladesh.* Dhaka: Dhaka Transport Coordination Authority (DTCA).

DTCA. 2016. *The Project on the Revision and Updating of the Strategic Transport Plan for Dhaka.* Dhaka Transport Coordination Authority (DTCA) and Japan International Cooperation Agency (JICA).

DTCB. 2005. *Strategic Transport Plan (STP) for Dhaka, Final Report.* Dhaka Transport Coordination Board (DTCB), Dhaka, Bangladesh.

Duarte, F. and Rojas, F. 2012. Intermodal connectivity to BRT: a comparative analysis of Bogotá and Curitiba. *Journal of Public Transportation.* **15**(2), p1.

Enam, A. 2010. *Developing a comprehensive mode choice model to capture the preferences for mass rapid transit in Dhaka.* . MSc. thesis, Bangladesh Universiy of Engineering and Technology.

Ettema, D., Furtado, B., Sakowski, P. and Tovolli, M. 2014. Complexity methods applied to transport planning. *Furtado, BA; Sakowski, PAM & Tovolli, MH.* pp.279-300.

Forward, S.E. 2019. Views on public transport and how personal experiences can contribute to a more positive attitude and behavioural change. *Social Sciences.* **8**(2), p47.

Gao, M., Zhou, L. and Chen, Y. 2016. An alternative approach for high speed railway carrying capacity calculation based on multiagent simulation. *Discrete Dynamics in Nature and Society.* **2016**.

Gkiotsalitis, K. and Stathopoulos, A. 2015. A utility-maximization model for retrieving users' willingness to travel for participating in activities from big-data. *Transportation Research Part C: Emerging Technologies.* **58**, pp.265-277.

Grether, D., Chen, Y., Rieser, M. and Nagel, K. 2008. Integrating a simple model for mode choice into a multi-agent simulation of travel behavior and traffic flow.

Habib, K.M.N. 2002. *Evaluation of Planning Options to Alleviate Traffic Congestion and Resulting Air Pollution in Dhaka City.* . MSc. thesis, Bangladesh University of Engineering and Technology.

Hafezi, M.H., Liu, L. and Millward, H. 2019. A time-use activity-pattern recognition model for activity-based travel demand modeling. *Transportation.* **46**(4), pp.1369-1394.

Han, B., Kim, J. and Timmermans, H. 2019. Task allocation and gender roles in dual earner households: the issue of escorting children. *Travel behaviour and society.* **14**, pp.11-20.

Hess, S. and Palma, D. 2019. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of choice modelling.* **32**, p100170.

Hidalgo, D. and Graftieaux, P. 2008. Bus rapid transit systems in Latin America and Asia: results and difficulties in 11 cities. *Transportation Research Record.* **2072**(1), pp.77-88.

Huang, R. 2019. Simulating individual work trips for transit-facilitated accessibility study. *Environment and Planning B: Urban Analytics and City Science.* **46**(1), pp.84-102.

Ji, Y., Gao, L., Chen, D., Ma, X. and Zhang, R. 2018. How does a static measure influence passengers' boarding behaviors and bus dwell time? Simulated evidence from Nanjing bus stations. *Transportation Research Part A: Policy and Practice.* **110**, pp.13-25.

Jing, P., You, Q. and Chen, L. 2018. Agent-Based Simulation of Children's School Travel Mode with Parental Escort Decisions. *Information.* **9**(3), p50.

Joseph, L., Neven, A., Martens, K., Kweka, O., Wets, G. and Janssens, D. 2021. Exploring changes in mobility experiences and perceptions after implementation of the bus rapid transit system in Dar es Salaam. *Case Studies on Transport Policy.* **9**(2), pp.930-938.

Kadiyali, L.R. 2010. *Traffic Engineering and Transport Planning.* Delhi, India: Khanna Publishers

Kagho, G.O., Balac, M. and Axhausen, K.W. 2020. Agent-based models in transport planning: Current state, issues, and expectations. *Procedia Computer Science.* **170**, pp.726-732.

Levine, J., Zellner, M., Arquero de Alarcón, M., Shiftan, Y. and Massey, D. 2018. The impact of automated transit, pedestrian, and bicycling facilities on urban travel patterns. *Transportation planning and technology.* **41**(5), pp.463-480.

Li, X.-y., Li, X.-m., Yang, L. and Li, J. 2018. Dynamic route and departure time choice model based on self-adaptive reference point and reinforcement learning. *Physica A: Statistical Mechanics and its Applications.* **502**, pp.77-92.

Madlener, R. and Sunak, Y. 2011. Impacts of urbanization on urban structures and energy demand: What can we learn for urban energy planning and urbanization management? *Sustainable Cities and Society.* **1**(1), pp.45-53.

Makinde, O., Neagu, D. and Gheorghe, M. 2018. Agent based micro-simulation of a passenger rail system using customer survey data and an activity based approach. In: *UK Workshop on Computational Intelligence*: Springer, pp.123-137.

Małecki, K. 2018. A computer simulation of traffic flow with on-street parking and drivers' behaviour based on cellular automata and a multi-agent system. *Journal of computational science.* **28**, pp.32-42.

Manley, E., Cheng, T., Penn, A. and Emmonds, A. 2014. A framework for simulating large-scale complex urban traffic dynamics through hybrid agent-based modelling. *Computers, Environment and Urban Systems.* **44**, pp.27-36.

Manser, P., Becker, H., Hörl, S. and Axhausen, K.W. 2020. Designing a large-scale public transport network using agent-based microsimulation. *Transportation Research Part A: Policy and Practice.* **137**, pp.1-15.

Märki, F., Charypar, D. and Axhausen, K.W. 2014. Agent-based model for continuous activity planning with an open planning horizon. *Transportation.* **41**(4), pp.905-922.

Mavi, R.K., Zarbakhshnia, N. and Khazraei, A. 2018. Bus rapid transit (BRT): A simulation and multi criteria decision making (MCDM) approach. *Transport Policy.* **72**, pp.187-197.

McCormack, G.R., Ghoneim, D., Frehlich, L., Blackstaffe, A., Turley, L. and Bracic, B. 2021. A 12-month natural experiment investigating the impacts of replacing a traditional bus service with bus rapid transit on physical activity. *Journal of Transport & Health.* **22**, p101239.

McDonnell, S. and Zellner, M. 2011. Exploring the effectiveness of bus rapid transit a prototype agent-based model of commuting behavior. *Transport Policy.* **18**(6), pp.825-835.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior.

Mehlstäubler, K. 2019. *Implementation of Public Transportation in the MATSim Simulation Environment*. Master's thesis, Technical University of Munich.

Melo, P.C., Graham, D.J. and Canavan, S. 2012. Effects of road investments on economic output and induced travel demand: evidence for urbanized areas in the United States. *Transportation research record.* **2297**(1), pp.163-171.

Moreno, A.T., Michalski, A., Llorca, C. and Moeckel, R. 2018. Shared autonomous vehicles effect on vehicle-km traveled and average trip duration. *Journal of Advanced Transportation.* **2018**.

Paget-Seekins, L. 2015. Bus rapid transit as a neoliberal contradiction. *Journal of Transport Geography.* **48**, pp.115-120.

Palacios, M.S., Cochran, A., Bell, C., Jiménez, U.H., Leshner, E., Morales, F.T. and Chatman, D.G. 2020. Bus rapid transit arrives in Barranquilla, Colombia: Understanding a changing landscape through residents' travel experiences. *Travel Behaviour and Society.* **21**, pp.131-139.

Poku-Boansi, M. and Marsden, G. 2018. Bus rapid transit systems as a governance reform project. *Journal of transport geography.* **70**, pp.193-202.

Quddus, M., Rahman, F., Monsuur, F., de Ona, J. and Enoch, M. 2019. Analyzing bus passengers' satisfaction in Dhaka using discrete choice models. *Transportation Research Record.* **2673**(2), pp.758-768.

Rahman, F. 2022. Exploring paratransit service quality based on low-income women's perception: A case study in Dhaka city by structural equation model (SEM). *IATSS Research.* **46**(2), pp.181-192.

Rahman, M.S.-U., Timms, P. and Montgomery, F. 2012. Integrating BRT systems with rickshaws in developing cities to promote energy efficient travel. *Procedia-Social and Behavioral Sciences.* **54**, pp.261-274.

Ramos, S., Vicente, P., Passos, A.M., Costa, P. and Reis, E. 2019. Perceptions of the public transport service as a barrier to the adoption of public transport: A qualitative study. *Social Sciences.* **8**(5), p150.

Raney, B., Cetin, N., Völlmy, A., Vrtic, M., Axhausen, K. and Nagel, K. 2003. An agent-based microsimulation model of Swiss travel: First results. *Networks and Spatial Economics.* **3**(1), pp.23-41.

Rashedi, Z., Mahmoud, M., Hasnine, S. and Habib, K.N. 2017. On the factors affecting the choice of regional transit for commuting in Greater Toronto and Hamilton Area: Application of an advanced RP-SP choice model. *Transportation Research Part A: Policy and Practice.* **105**, pp.1-13.

Rastogi, R. 2000. Stated Preference Survey: Experience From Developing Country. *Unpublished report, Department of Civil Engineering, Engineering College, KOTA Mumbai, India.*

Sajib, S.H. 2021. A Qualitative case study on inadequate public transport system and its social impact on people in Dhaka. *Transactions on Transport Sciences.* **12**(2).

Sanko, N. 2014. Travel demand forecasts improved by using cross-sectional data from multiple time points. *Transportation.* **41**(4), pp.673-695.

Satiennam, T., Jaensirisak, S., Satiennam, W. and Detdamrong, S. 2016. Potential for modal shift by passenger car and motorcycle users towards Bus Rapid Transit (BRT) in an Asian developing city. *IATSS research.* **39**(2), pp.121-129.

Satu, S.A. and Chiu, R.L. 2019. Livability in dense residential neighbourhoods of Dhaka. *Housing Studies.* **34**(3), pp.538-559.

Schalekamp, H. and Behrens, R. 2010. Engaging paratransit on public transport reform initiatives in South Africa: A critique of policy and an investigation of appropriate engagement approaches. *Research in Transportation Economics.* **29**(1), pp.371-378.

Shabanpour, R., Golshani, N., Auld, J. and Mohammadian, A. 2017. Dynamics of activity time-of-day choice. *Transportation Research Record.* **2665**(1), pp.51-59.

Shafi, R., Delbosc, A. and Rose, G. 2022. The role of culture and evolving attitudes in travel behaviour assimilation among south asian immigrants in Melbourne, Australia. *Transportation.* pp.1-27.

Shi, X., Moudon, A.V., Hurvitz, P.M., Mooney, S.J., Zhou, C. and Saelens, B.E. 2021. Does improving stop amenities help increase Bus Rapid Transit ridership? Findings based on a quasi-experiment. *Transportation research interdisciplinary perspectives.* **10**, p100323.

Siddique, M., Bakkar, A. and Choudhury, C.F. 2017. Modelling the behavioural response to congestion pricing in Dhaka, Bangladesh. *Transportation in Developing Economies.* **3**(2), pp.1-13.

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

TYPSA. 2019. *Feasibility study and preliminary design for construction of Dhaka subway.* Dhaka: TYPSA.

Västberg, O.B., Karlström, A., Jonsson, D. and Sundberg, M. 2020. A dynamic discrete choice activity-based travel demand model. *Transportation science.* **54**(1), pp.21-41.

Venter, C., Jennings, G., Hidalgo, D. and Valderrama Pineda, A.F. 2018. The equity impacts of bus rapid transit: A review of the evidence and implications for sustainable transport. *International Journal of Sustainable Transportation.* **12**(2), pp.140-152.

Vickrey, W.S. 1969. Congestion theory and transport investment. *The American Economic Review.* **59**(2), pp.251-260.
Vosooghi, R., Puchinger, J., Jankovic, M. and Sirin, G. 2017. A critical analysis of travel demand estimation for new one-way carsharing systems. In: *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*: IEEE, pp.199-205.
Wadud, Z. 2020. The effects of e-ridehailing on motorcycle ownership in an emerging-country megacity. *Transportation research part A: policy and practice.* **137**, pp.301-312.
WB. 2015. *Open Transport Development and Integration of ICT and Transport.* World Bank.
Wee, B. and Kroesen, M. 2022. oAttitude changes, modelling travel behaviour, and ex ante project evaluations. *Transportation Research Interdisciplinary Perspectives.* **16**, p100724.
Yagi, S. and Mohammadian, A. 2008. Policy simulation for new BRT and area pricing alternatives using an opinion survey in Jakarta. *Transportation Planning and Technology.* **31**(5), pp.589-612.
Yagi, S. and Mohammadian, A.K. 2010. An activity-based microsimulation model of travel demand in the Jakarta metropolitan area. *Journal of Choice Modelling.* **3**(1), pp.32-57.
Zgheib, N., Abou-Zeid, M. and Kaysi, I. 2020. Modeling demand for ridesourcing as feeder for high capacity mass transit systems with an application to the planned Beirut BRT. *Transportation Research Part A: Policy and Practice.* **138**, pp.70-91.
Zhang, L. and Levinson, D. 2004. Agent-based approach to travel demand modeling: Exploratory analysis. *Transportation Research Record.* **1898**(1), pp.28-36.
Zhang, S. and Zhao, J. 2018. Low-carbon futures for Shenzhen's urban passenger transport: A human-based approach. *Transportation Research Part D: Transport and Environment.* **62**, pp.236-255.
Zmud, J.P. and Sener, I.N. 2017. Towards an understanding of the travel behavior impact of autonomous vehicles. *Transportation research procedia.* **25**, pp.2500-2519.
Zou, M., Li, M., Lin, X., Xiong, C., Mao, C., Wan, C., Zhang, K. and Yu, J. 2016. An agent-based choice model for travel mode and departure time and its case study in Beijing. *Transportation Research Part C: Emerging Technologies.* **64**, pp.133-147.

Appendix D

All reproducible codes can be found in the following link: khzannat/dhaka-matsim-example-project (github.com). Activity plan file is not shared due to data privacy.

Supplementary Table 6-1 Distribution of trip purpose under different scenarios (scenario 2 and 3 is the distribution of home based and non-home-based BRT passenger trip purpose).

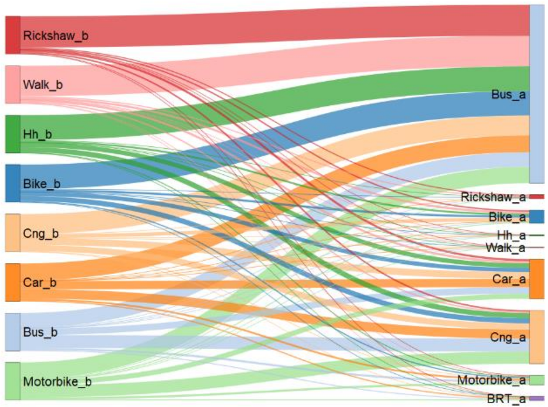
	Scenario 1		**Scenario 2A**		**Scenario 2B**		**Scenario 2C**		**Scenario 3A**		**Scenario 3B**	
	Home based	**Non home based**	**Home based**	**Non home based**	**Home based**	**Non home based**	**Home based**	**Non home based**	**Home based**	**Non home based**	**Home based**	**Non home based**
	95.2	4.8	100	0	91.5	8.5	90.1	8.90	93.6	6.4	93.7	6.3
Work	57.2	72.0	49.0	0	65.3	77.9	67.5	80.0	63.4	74.4	63.4	73.6
Education	25.8	3.7	49.0	0	21.0	2.9	20.4	1.2	21.9	2.0	21.8	2.1
Personal	9.3	14.9	0.0	0	9.1	9.6	8.3	8.5	9.6	14.6	9.6	15.2
Leisure	0.6	0.6	0.0	0	0.9	0.0	0.7	0.0	0.8	0.5	0.7	0.6
Shopping	4.0	5.6	2.0	0	2.5	8.1	2.1	7.9	2.9	5.6	2.9	5.8
Other	3.1	3.2	0.0	0	1.2	1.5	1.0	2.4	1.5	2.9	1.5	2.7

Supplementary Table 6-2 Distribution of departure time under different scenarios.

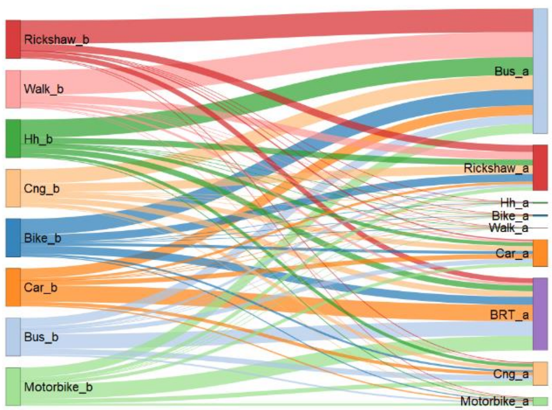
Departure time	Scenario 1	Scenario 2A	Scenario 2B	Scenario 2C	Scenario 3A	Scenario 3B
1	0.07	0.00	0.00	0.00	0.00	0.00
2	0.06	2.04	0.00	0.00	0.00	0.00
3	0.06	0.00	0.00	0.00	0.01	0.01
4	0.13	0.00	0.00	0.00	0.02	0.03
5	0.50	12.24	0.59	0.62	0.37	0.34
6	2.20	18.37	2.95	2.78	2.18	2.17
7	9.35	30.61	9.17	8.86	8.51	8.59
8	13.62	10.20	14.55	14.65	14.20	14.24
9	9.25	4.08	9.31	9.54	10.06	10.06
10	6.97	0.00	7.99	7.67	7.04	7.10
11	4.59	4.08	3.15	3.35	3.75	3.63
12	4.98	2.04	3.93	4.26	3.99	3.93
13	5.27	2.04	5.18	4.66	4.60	4.57
14	5.77	0.00	6.42	6.19	5.75	5.81
15	5.19	0.00	5.57	6.13	5.78	5.74
16	6.50	2.04	8.39	9.09	7.63	7.64
17	6.13	4.08	5.77	5.62	7.28	7.29
18	4.46	2.04	4.06	4.37	5.18	5.35
19	4.51	2.04	4.46	3.80	4.59	4.64
20	4.02	2.04	3.54	4.03	3.95	3.98
21	3.37	2.04	3.47	3.07	3.39	3.25
22	2.19	0.00	1.44	1.25	1.64	1.60
23	0.69	0.00	0.07	0.06	0.07	0.04
24	0.12	0.00	0.00	0.00	0.00	0.00

Supplementary Table 6-3 Socio-demographic profile of BRT users in different scenarios.

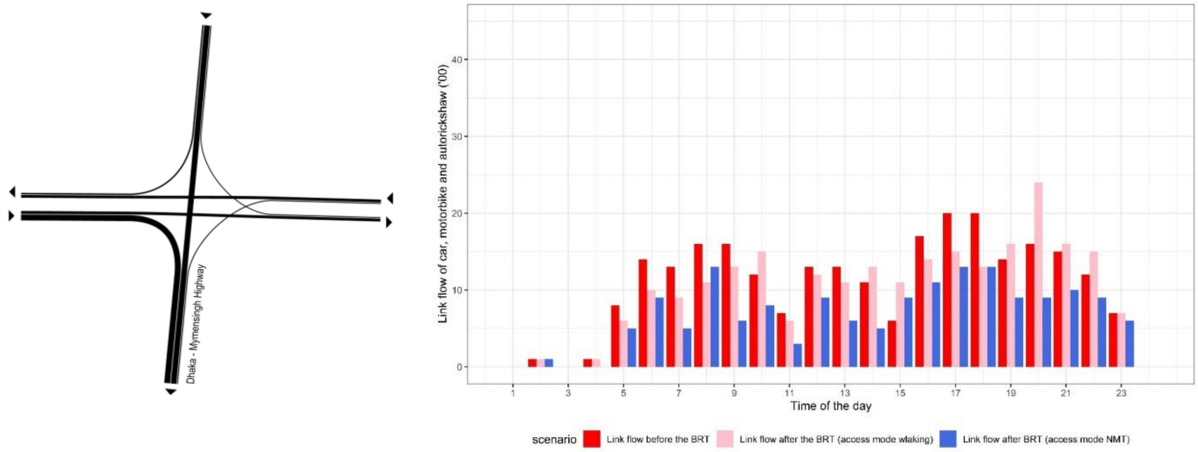
	Employed	Working age (24 to 40)	Male	Having car at the household
Proportion in the sample	60.95%	38.48%	73.71%	6.18%
Scenario 2B	73.3%	41.3%	79.6%	7.9%
Scenario 2C	73.7%	38.7%	81.0%	8.7%
Scenario 3A	67.1%	40.1%	81.5%	7.9%
Scenario 3B	68.2%	40.8%	81.7%	7.8%



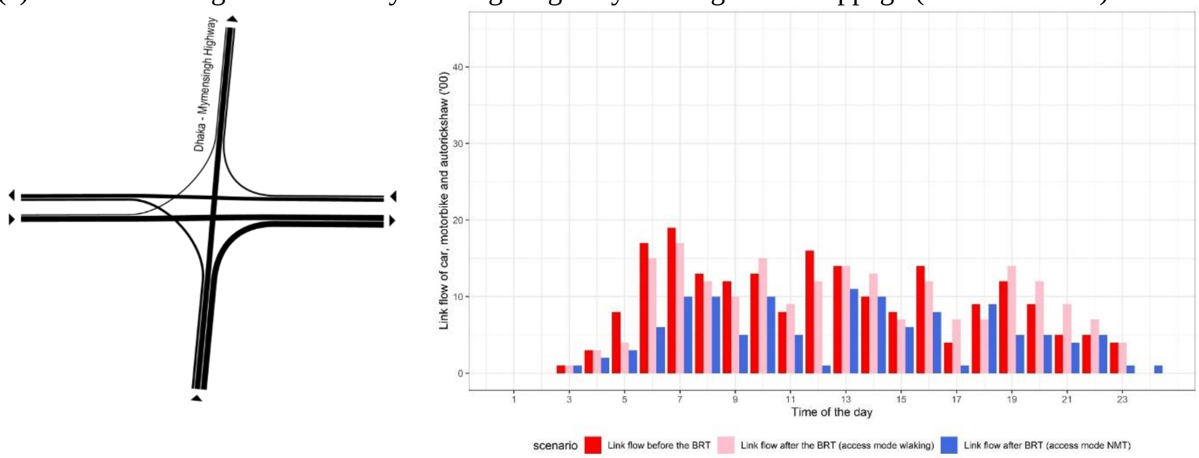
Supplementary Figure 6-1 Modal shift based on travel time-travel cost (monthly flat fare) without multi-modal accessibility.



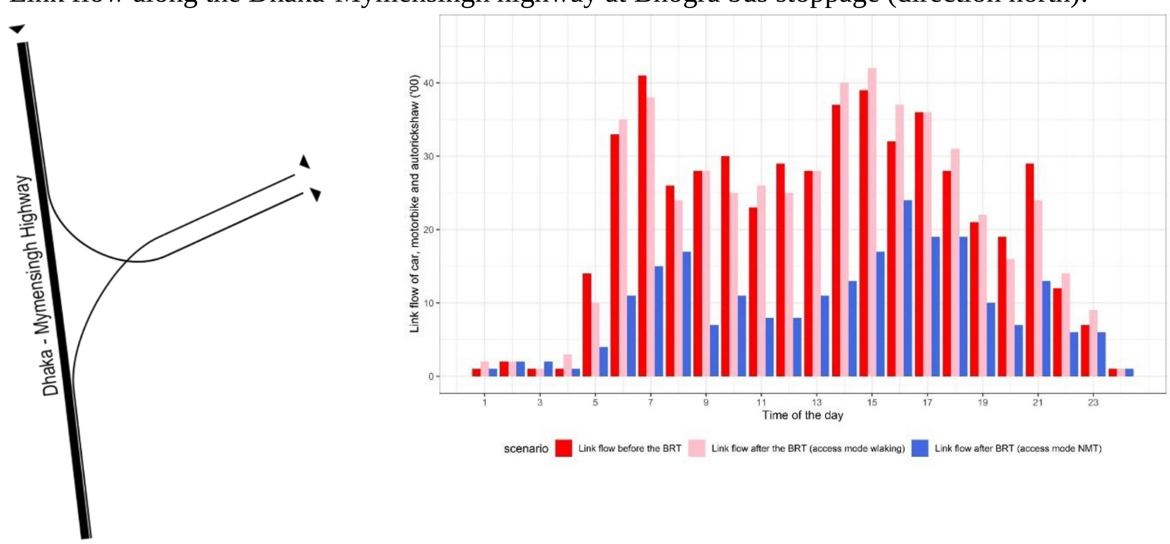
Supplementary Figure 6-2 Modal shift based on travel time-travel cost (monthly flat fare) with multi-modal access-egress facility (function of distance).



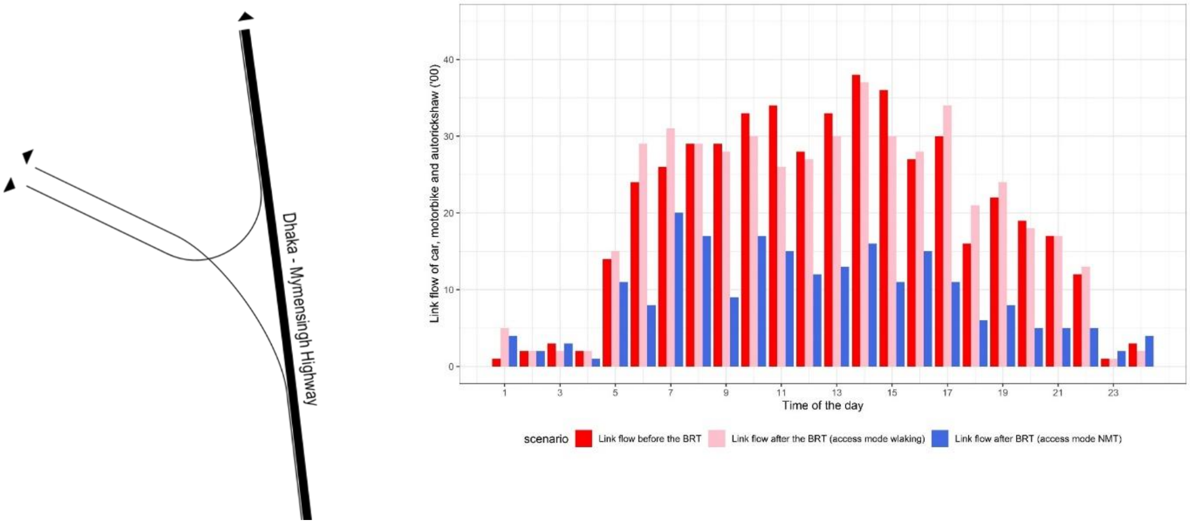
(a) Link flow along the Dhaka-Mymensingh highway at Bhogra bus stoppage (direction south).



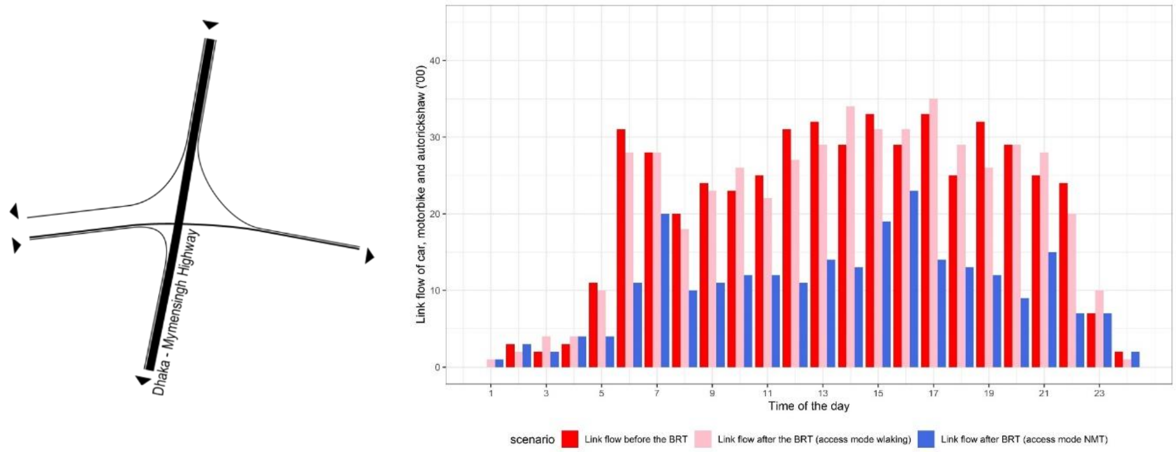
(b) Link flow along the Dhaka-Mymensingh highway at Bhogra bus stoppage (direction north).



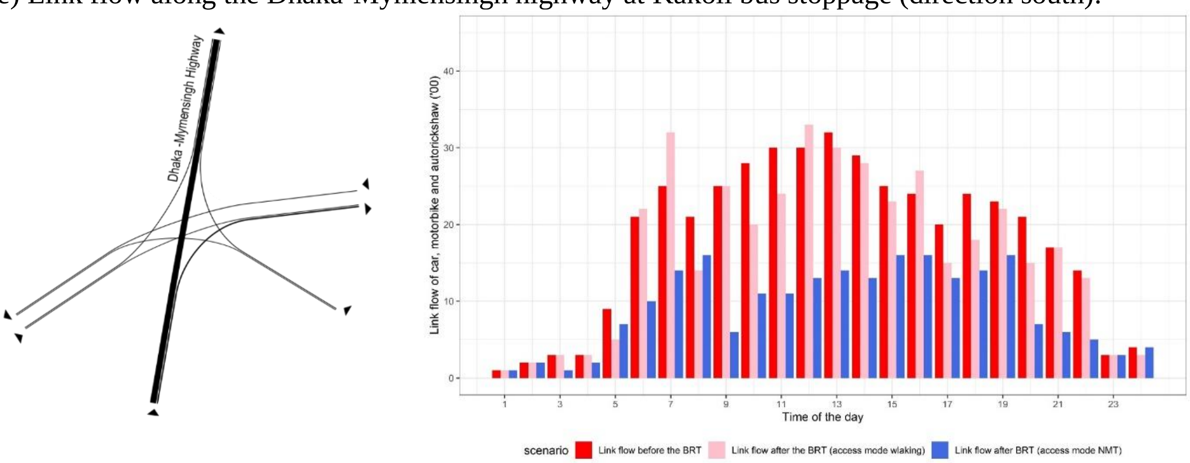
(c) Link flow along the Dhaka-Mymensingh highway at Khilkhet bus stoppage (direction south).



(d) Link flow along the Dhaka-Mymensingh highway at Khilkhet bus stoppage (direction north).



(e) Link flow along the Dhaka-Mymensingh highway at Kakoli bus stoppage (direction south).

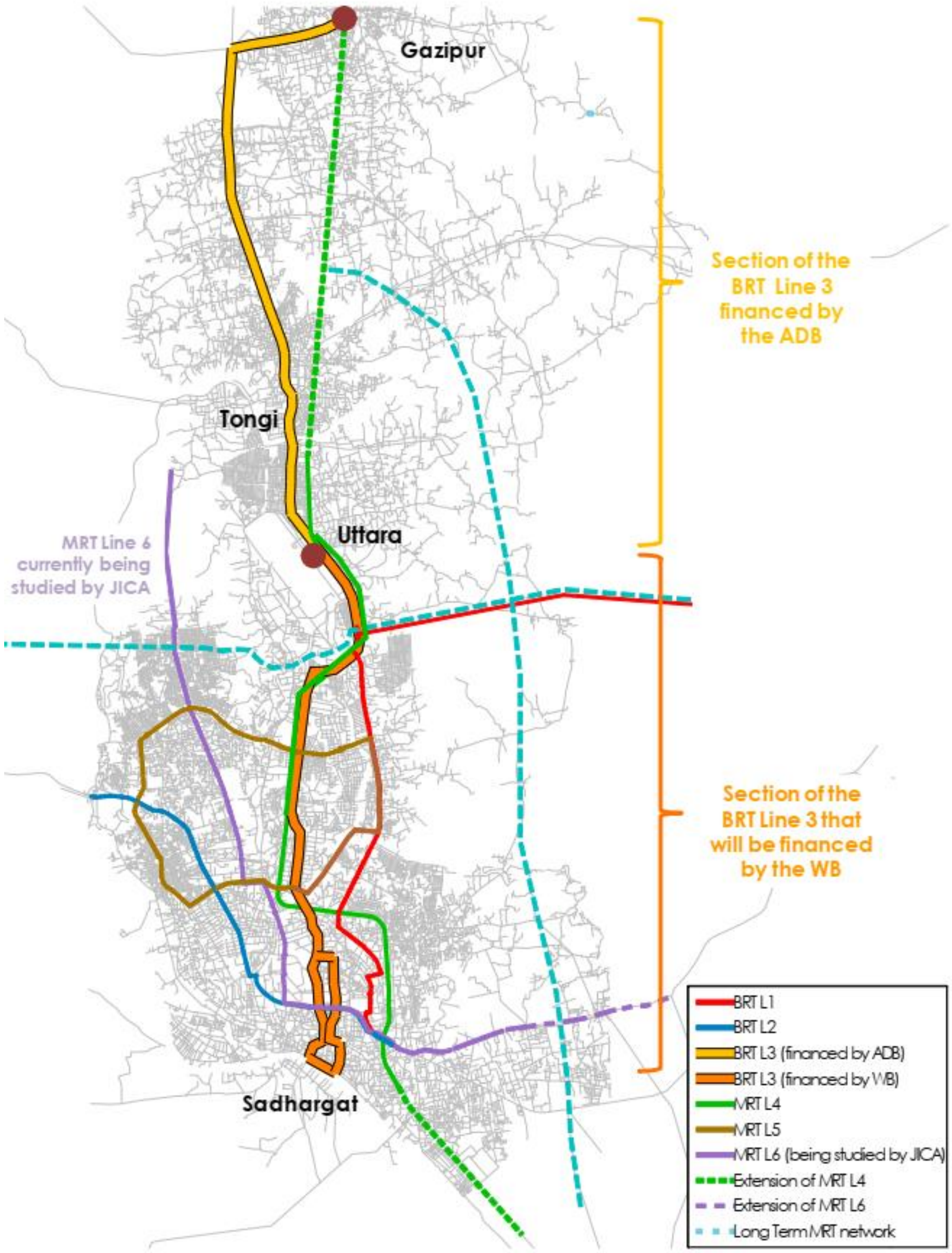


(f) Link flow along the Dhaka-Mymensingh highway at Kakoli bus stoppage (direction north).

Supplementary Figure 6-3 Link flow along the Dhaka-Mymensingh highway.

A= Bus, B=Auto-rickshaw, C= Car, D= Motorcycle, E= Bicycle, F=Rickshaw, G= Human-hauler, H= Pedestrian

Supplementary Figure 6-4 Passenger transport modes available in Dhaka city.



Supplementary Figure 6-5 Greater Dhaka Sustainable Urban Transport Corridor (GDSUTP) within the framework of the STP proposals.

Chapter 7 Investigating the relative precision of GPS, GSM and CDR data for inferring spatiotemporal travel trajectories

Abstract

Over the past decade, the potential of passively generated big data sources (such as Global Positioning System (GPS), Call Detail Records (CDR), and Global System for Mobile Communication (GSM)) in transport modelling and behavioural analysis has been firmly established. However, the challenge lies in evaluating the accuracy and usefulness of these passive data sources for transport policy and planning, primarily due to the absence of ground-truth (GT) data and the resulting dearth of validation studies. Some recent studies have attempted to address this challenge by using externally collected GT data (e.g., travel diary survey, census) or internally collected parallel samples (e.g., GPS or app-based location updates). Nevertheless, a notable research gap remains in dealing with the unique challenges associated with accurately assessing and validating mobility information extracted from passive data sources with appropriate GT that precisely matches the spatial and temporal granularity of these passive data sources. The aim of this study is to assess the accuracy of inferring human mobility patterns from three commonly used passively generated data sources: anonymous GPS traces, CDR, and GSM from mobile phone. In the absence of real-world disaggregate ground truth data for benchmarking, we used outputs from an agent-based simulation platform (MATSim simulation) as 'synthetic ground truth' (SGT) for each traveller (agent). Subsequently, for each agent, we generated synthetic GPS, GSM, and CDR data corresponding to the SGT, while considering their positional disturbances and utilising the conventional spatial and temporal resolution specific to each type of passive data. We extracted mobility information by identifying potential stay locations, which enabled us to determine home and activity locations. This information was then used to measure trip-related attributes such as departure time and trajectory distance. The statistical and spatial characteristics of individual mobility information obtained from the GPS, GSM, and CDR data were compared with the SGT data to evaluate the accuracy of passive trajectory data at both the disaggregate and aggregate levels. The results provided insights into the accuracy of outputs derived from these data sources and the degree of bias in comparison to the SGT. The findings are expected to be useful for transport researchers and practitioners in selecting the most suitable data type for specific analyses and understanding the potential margin of error involved.

Keywords: GPS, GSM, CDR, agent-based simulation, MATSim, synthetic data

7.1. Introduction

Over the past decade, passively generated spatiotemporal data have emerged as popular sources for extracting mobility information, such as activity location, departure time, and mode of transportation. Among various types of passive data, the most commonly used data for understanding travel behaviour and travel demand modelling include anonymous Global Positioning System (GPS), Call Detail Records (CDR), Global System for Mobile communication (GSM) data, social media data, and public transport smart card records (Pelletier et al., 2011; Shen and Stopher, 2014; Wang et al., 2018; Huang et al., 2019). These datasets are typically (Harrison et al., 2020) characterised by their large size and the provision of updated, near-real-time spatial and temporal information from a substantial sample size over an extended observation period (Zannat and Choudhury, 2019). However, utilising such data for transport planning poses challenges primarily due to their varying strengths and weaknesses (Supplementary Table 7-1), restrictions imposed by General Data Protection Regulation (GDPR)[29], and the presence of various types of noise that impact the accuracy of extracted mobility information (Liu et al., 2016). Errors in data can arise from different sources, including the devices and technology used for data collection, data processing skills, software, and algorithms employed (Eagle and Pentland, 2006). For example, multipath interference and satellite errors in GPS data, tower location density and cell size effect in GSM and CDR data (Forghani et al., 2020; Hendawi et al., 2020). Additionally, noise may appear due to aspects related to users, such as randomness in users' mobility behaviour, access to the service (which could be influenced by network coverage and socio-demographic characteristics), communication, and technology usage patterns (Etter et al., 2012; Wu et al., 2018; Dong, 2022). For example, the frequency of phone calls and, the number of users using a single device influence the quality of CDR data (Bwambale, A. et al., 2019; Li et al., 2023). Other contextual factors such as land use, topography, vegetation, and urban density can introduce additional inaccuracies in the data (Ahas et al., 2007b; Paul and Rimer, 2013; Chen, C. et al., 2016; Fund et al., 2016; Anda et al., 2017; Kang et al., 2020). Therefore, it is essential to assess the reliability and precision of trajectory information extracted from various passive sources before employing mobility information for transport planning and modelling. Such accuracy assessment to validate the passive trajectory data as an alternative to conventional survey data is particularly important in the context of the Global South where the collection of survey data is often labour and resource intensive.

In accordance with various types of noise, many studies have concentrated on the accuracy and precision of passive trajectory data and its determinants, delineating three principal perspectives: data source (technology), spatial context, and user attributes. Evaluating accuracy from the standpoint of data source and spatial context underscores the crucial significance of temporal and spatial precision within trajectory data, as these factors directly impact the validity of research on human mobility (Ahas et al., 2007a; Song et al., 2020). For instance, when utilising passive data for trip-based demand modelling as the starting point for generating origin-destination (O-D) matrices, precise identification of the geo-referenced location and trip timing is a prerequisite (Iqbal et al., 2014). Similarly, to use passive data in activity-and

[29] The implementation of GDPR varies significantly across different regions globally, and its presence can impact data quality and availability (von Mörner, 2017; Jansen et al., 2021). For example, it might result in the exclusion of brief and infrequent travel patterns, which could potentially expose personally identifiable information. In this particular context, the GDPR did not affect the results as it is not in place in Dhaka (and in many other parts of the developing world).

agent-based demand modelling, precise information regarding activity locations is necessary to infer trip purposes and accurately reflect corresponding travel patterns (Liao et al., 2007). While methodologies for assessing the accuracy of trajectory data collected by survey (e.g., manual survey, smartphone app-based survey) and validating them against ground truth (GT) are well-established (Jones and Stopher, 2003; Wolf et al., 2004; Stopher et al., 2007; Kohla and Meschik, 2013; Zmud et al., 2013; Janzen et al., 2018; Harding et al., 2021), there remains a significant research gap in addressing the specific challenges of accurately assessing and validating mobility information derived from anonymous passive data sources (Harding et al., 2021). One reason for the scarcity of accuracy assessment (both positional and temporal) studies on passive trajectory data is the limited availability of suitable GT that matches spatial and temporal granularity for these data sources.

In the contemporary literature, GT data used to validate passive trajectory data can be broadly categorised into two classes: (1) externally collected data, including census data, travel surveys and traffic counts, and (2) internally collected data, which includes information concurrently recorded with passive data through parallel surveys involving subsets of individuals using GPS or app-based location updates (Bar-Gera, 2007; Bohte and Maat, 2009; Sternfeld et al., 2012; Tsoleridis et al., 2022; Pan et al., 2023). Different studies have attempted to validate passive trajectory data against externally collected GT data. For instance, Vanhoof et al. (2020) compared location information extracted from mobile phone data with census data. Gordon et al. (2013) manually collected trace counts to compare boarding and alighting times, locations, and interchanges inferred from automatic fare collection (AFC) and automatic vehicle location (AVL) data. By contrast, a few studies have validated passive trajectory data using internally collected data, where the actual location (e.g., home or activity location) of a subset of individuals was known in advance due to their voluntary participation as mobile phone users (Pappalardo et al., 2020; Song et al., 2020; Yang et al., 2021). Bwambale, A. et al. (2019) compared travel time sensitivity, schedule delay, and stoppage number extracted using two different data types with different temporal resolutions (GPS and GSM) in the context of departure time choice. In an experimental study, Forghani et al. (2020) compared trajectory generated from CDR data with GPS logger. Some studies have also used both externally and internally collected data simultaneously to validate passive data. For instance, Toole et al. (2015) leveraged mobile phone call data records to extract both the Origin-Destination (O-D) matrix and routes, incorporating census data and travel diary survey information.

Evaluating the spatiotemporal precision of passive trajectory data against real-world GT data that accurately reflects the true trajectories of respondents is a formidable challenge for several reasons. Validating against externally collected data often requires spatial and temporal adjustments to make meaningful comparisons with passive data. For instance, conventional travel surveys often provide location and time information at the traffic analysis zone level (TAZ), while passive trajectory data (e.g., mobile phone data, smart card data, etc.) offer individual-level location updates throughout a journey. Additionally, census or travel data may lack the necessary information for modelling, such as departure times, short stays, or delays due to traffic (Hato, 2010; Oliveira et al., 2011). Consequently, when models developed with passive data are validated against traditional traffic counts or travel surveys, there is a risk of overestimating or underestimating predicted travel demand. For instance, short trips are frequently omitted from travel surveys, and models built using GPS data may be labelled as overestimating travel demand because GPS data can capture information about short stays and visits. Conversely, parallel surveys (i.e., internally collected data) involving GPS or apps for individuals can be more complex, costly (particularly for developing countries), and less feasible for extensive passive datasets like GSM and CDR records (Pan et al., 2023). Furthermore, passive data acquired through third-party entities, such as mobile network operators or service providers, are subject to privacy concerns and data protection regulations. Consequently, access to passive data is typically granted to transport modellers in an aggregated and anonymised format. This anonymisation process complicates the attribution of devices or services to specific individuals, making it challenging to collect true travel trajectories associated with a given device or service.

Many challenges related to externally and internally collected GT data could be addressed by generating, synthetic ground truth (SGT) data compatible with different passive data types and their corresponding spatial and temporal resolutions. SGT data offers distinct advantages by allowing the formation of a comprehensive range of potential scenarios, including different activity patterns, mode use, and activity locations, spanning a range of spatial and temporal resolutions depending on the type of passive data under consideration. This capability is lacking in both externally and internally collected GT data. Synthetic data is often used for accuracy assessment when it is difficult to access real-world data or when true data is unavailable (Zilske and Nagel, 2014; Zhang et al., 2020). Nevertheless, similar examination of relative precision of spatiotemporal data extracted from passive data sources with appropriate (S)GT is still in its early stages. Further investigation of positional and temporal accuracy assessment with reliable GT is needed to make informed decisions based on passive data (Hemmings and Goves, 2017).

In this study, we proposed a framework for assessing the positional and temporal accuracy of passive trajectory data with appropriate GT that matches the spatiotemporal resolution of passive data. We considered three mainstream passive data types — anonymous data from GPS[30] device, GSM[31], and CDR[32] —for accuracy assessment. In absence of real-world disaggregate ground truth data for benchmarking, we treated outputs generated from an agent-based simulation platform (MATSim simulation) as the 'synthetic ground-truth' (SGT) for each traveller. Synthetic GPS, GSM and CDR data corresponding to this SGT were generated based on the standard spatial and temporal resolutions typical for each type of data. We compared the statistical and spatial characteristics of individual mobility information extracted from the GPS, GSM, and CDR data with the SGT data to assess the (temporal and positional) accuracy of passive trajectory data at both the disaggregate and aggregate levels. Another contribution of this work is the generation of a synthetic dataset, allowing us to quantify the magnitude of errors (due to the added noise) in different travel information extracted from passive data compared to GT. It is important to note that the current study is limited only to examining a straightforward type of noise i.e., spatial-temporal disturbance available in each source type, which impacts the positional and temporal accuracy of trajectory data. However, our proposed framework can be employed recursively to incorporate any inherent noise associated with the data under consideration or arising from different

[30] GPS data is timestamp location data generated by global positioning system satellites. These satellites emit signals that are received by GPS receivers in devices like smartphones, GPS trackers, or navigation device.

[31] GSM is a type of mobile phone data that provides triangulated location information when the mobile phone with a valid sim card is turned on. It offers higher temporal and spatial resolution than CDR data (another type of mobile phone data) as one single call will generate multiple sighting in GSM data.

[32] CDR data is a type of mobile phone data that records locations at the tower level when the mobile phone is in use (such as during calling and texting). In CDR data, each phone call generates a single record, representing the location associated with the tower used during that call.

geographical landscape which will enable the evaluation of positional and temporal accuracy and reliability in comparison to the ground truth generated.

The remainder of the paper is organised as follows: the following section discusses the methodology employed, as well as data sources used in this study. Subsequent sections present the analysis results, followed by a discussion of the findings. The paper concludes with a summary and directions for future research.

7.2. Methodology

The methodology employed in this article can be categorised into four main steps — 1) generating synthetic ground truth (SGT) data, 2) generating corresponding passive mobility data, 3) Extracting mobility attributes from the synthetic trajectory data, and 4) accuracy assessment. The overall methodology of this study is illustrated in Figure 7-1.

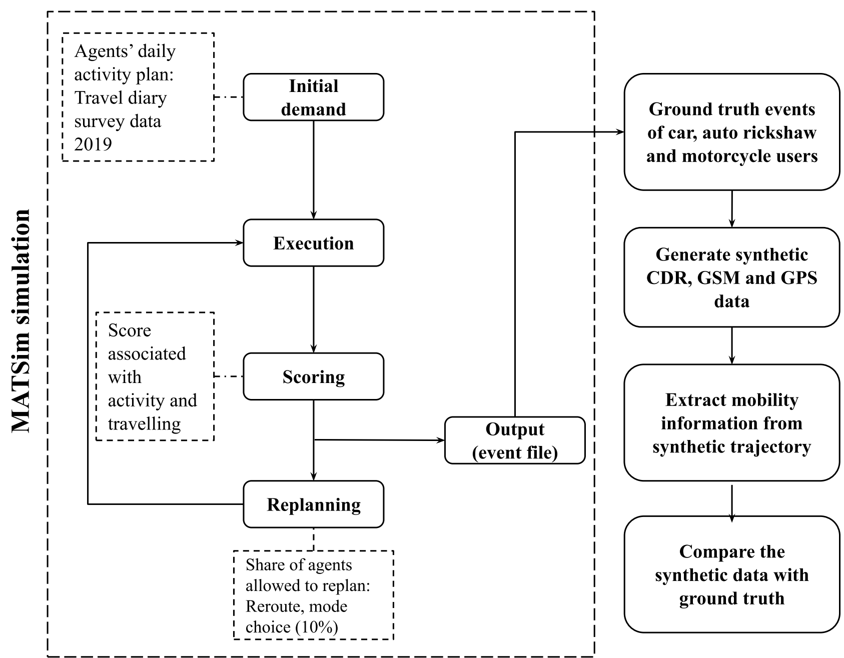


Figure 7-1 Framework to check the accuracy of trajectory data.

7.2.1. Generating synthetic ground truth (SGT)

The agent-based simulation tool MATSim (Multi-agent Transport Simulation) was used to generate the SGT data. At its core, MATSim operates by allowing a group of agents to interact within a virtual environment. The inputs required for MATSim simulation include activity plans, transport networks, and configuration files. An activity plan serves as a sequence of actions that agents are required to perform within this artificial setting. Typically, such plans are generated using microdata from a representative sample or synthetic population. In this study, we employed household-level trip diary data derived from the subway study by TYPSA (https://www.typsa.com/en/) in Dhaka, detailed demographic information for which can be found in Table 6-1.

However, the survey data could not be directly employed as ground truth because it lacked specific route information. In contrast, passive data typically contains location records within the trajectory while traversing. Furthermore, the trip diary data only offered activity locations at the TAZ level and provided detailed geographic information solely for each participant's home location. To address this, ArcGIS 10.8 was employed to randomly allocate activity locations within TAZ boundaries, ensuring compliance with the user-stated travel time (Bekhor et al., 2010). Following the assignment of activity locations, an activity profile was generated for each agent based on trip information extracted from the travel diary survey data. Each activity plan included information regarding the activity's location (x-y-coordinate), the end time of the first activity, the chosen 'leg' mode, and the maximum duration allocated for that activity.

To represent a virtual urban transport landscape mirroring Dhaka, this study integrated a road network and available information on existing transport services. The road network for the study area was obtained from the Open Street Map (OSM). In MATSim, the available modes include car, public transport (PT), bike, and walking. While Dhaka also features modes such as auto-rickshaws and motorbikes, these are not predefined options in MATSim. To account for them, they were modelled using special *vehicular specifications* within the existing framework of MATSim. The default configuration settings of the MATSim simulation were used to generate the SGT data.

In the MATSim simulation, each agent strives to optimise their actions based on a utility function. During the iterative process, a specific proportion of agents are allowed to modify their typical choices in an attempt to identify strategies with higher utility. This iterative process continues until the overall score of the population reaches equilibrium within the simulation. The strategy adopted by each agent at this equilibrium is intended to be a realistic approximation of their actual behaviour.

To generate the SGT data in this study, the marginal utility of travel time and cost derived from a joint RP-SP mode choice model was employed. Table 6-1 and Table 6-3 provides further details on the RP-SP data and the mode choice model used in the simulation respectively. A predetermined proportion of agents were allowed to change their trip mode (randomly selecting a leg mode) and route during the iteration process in an effort to find strategies with higher utility. All agents made attempts to adjust their plans to increase their utility by tracking each action chain. The simulation was run until the population achieved an equilibrium state, requiring a total of 800 iterations. Figure 7-2 presents the mode statistics of the simulation model. We compared the predicted modal share and preferred departure time with actual data collected from the 2019 travel diary survey. Additionally, the modal share of hypothetical passenger trips was compared with the modal share of the passenger trips collected by the Japan International Cooperation Agency (JICA) research team using the inner cordon line survey in Dhaka in 2014, details of which can be found in Table 6-6.

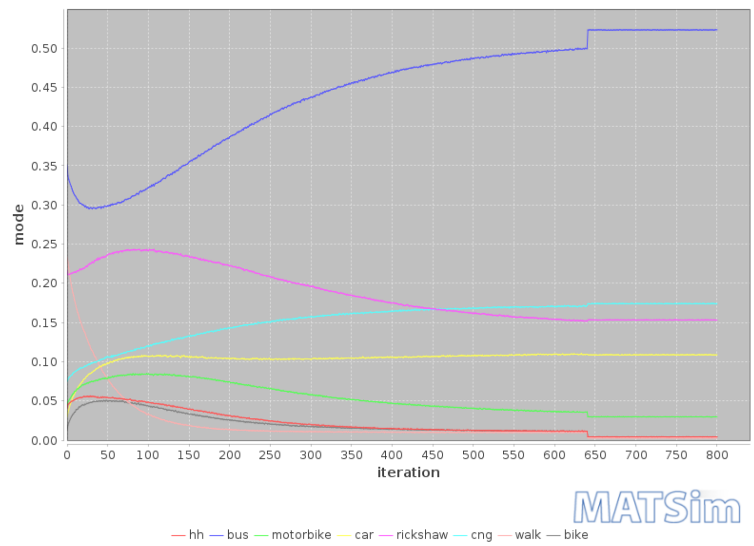


Figure 7-2 Mode statistics during the MATSim simulation.

The traffic simulation component was used to ascertain the "real" travel schedules for cars, motorcycles, and autorickshaws. As for the bus and human hauler routes, they were simulated utilising the "teleportation" feature, but with network-derived journey distances since precise information about these routes and schedules was lacking. Similarly, non-motorised modes were also simulated using the "teleportation" feature. Therefore, the simulation produced event files that comprehensively documented every specific action taken by agents who exclusively used motorcycles, autorickshaws, and cars throughout the simulation period. These actions included activity start times, activity end times, and link level interactions.

We extracted the complete, unprocessed list of actions carried out by agents who opted for a car, autorickshaw, or motorbike during the simulation period from the final event file. This extraction detailed the entire trajectory and activities of each agent. In this study, the extracted event file served as the SGT and the baseline for generating synthetic CDR, GSM, and GPS data. Notably, since PT users were simulated using the teleportation feature of MATSim, only the origin and destination, along with the activity start and finish times, were recorded in their event files, as opposed to their full trajectories. As a result, we limited our analyses to private mode users and excluded the PT users from the comparison.

7.2.2. Generating detailed trajectory data

Accessing real-world mobility-related passive data can be a challenging endeavour for transport modellers due to a myriad of concerns including privacy issues, the potential for re-identification, legal restrictions, data ownership, and data availability. To address these challenges, significant research efforts have focused on the generation of synthetic data. Such efforts aim to obfuscate or mask real-world location data derived from mobile phones and navigation devices for privacy and security purposes (Smith et al., 2009; Zilske and Nagel, 2014). In this study, we adopted the fundamental principles established in previous research within this domain. These principles underpin the following assumptions:

- It was assumed that the mobile device remained active throughout the entire journey and was carried by the travellers for the duration of their trips.

- All agents were presumed to choose the shortest route during their travels.
- The sociodemographic characteristics of the chosen agents were consistent.
- The mobile phone tower network strength within the study area was assumed to be uniform, with no differentiation between 2G, 3G, and 4G networks. Additionally, there were no gaps or incompleteness in the data due to topographic features or 'urban tunnelling' effects.

7.2.2.1. Synthetic GPS data

GPS data is generated by smartphones or GPS devices, which may exhibit varying degrees of measurement noise. These data typically consist of anonymous timestamped latitudes and longitudes. For this research, we generated synthetic data from GPS device. In Dhaka city, where there is no underground or tunnel infrastructure, GPS devices can provide relatively accurate location data when used aboveground. These devices record positions both outdoors and inside various structures such as buildings, buses, elevated trains, bridges, and within urban canyons (Gong et al., 2012). However, it is important to note that the accuracy of the recorded positions may vary. The positioning errors in GPS data from devices are influenced by two main factors: the type and quality of GPS receiver used, and the type of urban environment (atmospheric conditions, satellite signals, built environment), which can lead to poor-quality measurements and weaker satellite geometry. To generate synthetic GPS data, we followed a systematic procedure. The MATSim simulation yielded temporal data whenever an agent interacted with a specific event. In contrast, depending on the type of GPS receiver and signal strength in the target region, GPS devices provide more frequent location and time information. Therefore, to produce intermediary points that bridge the gap between the SGT point feature and create more realistic GPS data, we used linear interpolation. GPS points were generated at 10-second intervals. Typically, the positioning error for GPS-enabled systems falls within the range of 1 metre to 20 metres (Quddus et al., 2005; Hasan et al., 2009; Merry and Bettinger, 2019). To account for the geo-positioning noise, we introduced two types of simulated noise as horizontal positioning disturbance — random shift and random drift, following the method proposed by Bösche et al. (2013).

First, we applied Gaussian noise to the spatial data with a mean of 0 and a standard deviation of 3 metres. A well-designed GPS receiver generally offers horizontal accuracy of at least 3 metres (Renfro et al., 2022). However, this could lead to abrupt changes in direction between successive GPS points. To mitigate this, we applied shifts with a probability of 0.05 per second and otherwise, added the deviation from the previous point to the current point. Second, we simulated random drift with a probability of 0.03 per second, representing shifts perpendicular to the current driving direction. The maximum drift distance was determined from a Gaussian distribution with a mean of 0 and a standard deviation of 10 metres, as GPS in moving vehicles can achieve dynamic accuracy of up to 10 metres in urban areas (Schipperijn et al., 2014). It took precisely 30 seconds to reach the maximum drift distance, after which the drift distance gradually decreased to zero over another 30 seconds. Table 7-1 summarises all the parameters selected for adding noise to the GPS data.

Table 7-1 Summary of parameters used for adding noise while generating synthetic GPS and GSM data.

	Parameter	Value
GPS	Shift parameter (mean)	0
	Shift parameter (standard deviation)	3 metres
	Shift probability	0.05
	Drift parameter (mean)	0
	Drift parameter (standard deviation)	10 metres
	Drift Probability	0.03
GSM	Shift parameter (mean)	0
	Shift parameter (standard deviation)	5 metres
	Drift parameter (mean)	0
	Drift parameter (standard deviation)	15 metres
	Drift Probability	0.03

7.2.2.2. Synthetic GSM data

GSM data comprises the identifiers of all GSM cells that a mobile phone passes through at regular intervals during its use (Steenbruggen et al., 2013). Notably, GSM data offers finer spatial and temporal resolution compared to CDR data because it becomes accessible as soon as the phone, equipped with a valid SIM card, is activated. In our study, we leveraged the location data from the SGT to create synthetic GSM data. The temporal resolution of the GSM data in our study was set at 60 seconds, aligning with the temporal resolution of real GSM data (Bwambale, A. et al., 2019). For each sighting time, we selected locations from the respondent's SGT data while preserving the timestamped location sequence. If the time interval between two successive SGT points exceeded 60 seconds, we applied the linear interpolation method to generate intermediary cell points. However, it is crucial to acknowledge that the positioning accuracy of GSM data can widely, ranging from 1 metre to 600 metres, depending on factors such as the location type (indoor/outdoor), cell size and data collection techniques (triangulation, radio camera, signalling messages, GPS, etc.) (Chen, M.Y. et al., 2006; Ratti et al., 2006; Varshavsky et al., 2007; Steenbruggen et al., 2013). To generate synthetic GSM data for the densely urbanised and heavily populated area of Dhaka city, we added noise to the location points generated from SGT, which was relatively larger than the noise added with GPS data (GSM data tends to be noisier compared to GPS data, as per Bwambale, A. et al. (2019)). The summary of parameters used for generating synthetic GSM is outlined in Table 7-1. Gaussian noise (with a mean of 0 and stand deviation of 5 metres) was included as positioning disturbance. Given that successive GSM points can exhibit abrupt jumps from one side of the road to the other, we did not apply corrections for directional deviations in GSM points. However, similar to GPS data, we introduced random drift for moving vehicles at a probability rate of 0.03 per second in the case of GSM data.

7.2.2.3. Synthetic CDR data

Real-world CDR data includes time-stamped tower locations whenever a user initiates a phone call, sends a text message, or accesses mobile Wi-Fi. Table 7-2 provides an example of CDR data from various hypothetical users. In Dhaka, the most recent available CDR data was collected between June 19, 2012, and July 18, 2012. For detailed statistics regarding the available CDR data, refer to (Iqbal et al., 2014; Bwambale, A. et al., 2019).

Table 7-2 An excerpt from CDR data in a typical working day.

Unique caller ID	Date	Time	Call duration	Latitude	Longitude
AAH03JABkAAHvEkAQE	20120622	13:32:38	530	23.7186	90.4494
AAH03JACKAAAgfBALW	20120622	13:41:25	15	23.9139	90.2931
AAH03JAC8AAAbZfAHB	20120622	13:41:49	73	23.7911	90.2603
AAH03JAC5AAAdAkAJZ	20120622	13:45:40	16	23.7172	90.3556
AAH03JAC3AAAdDZAEe	20120622	13:46:22	17	23.1581	90.4119

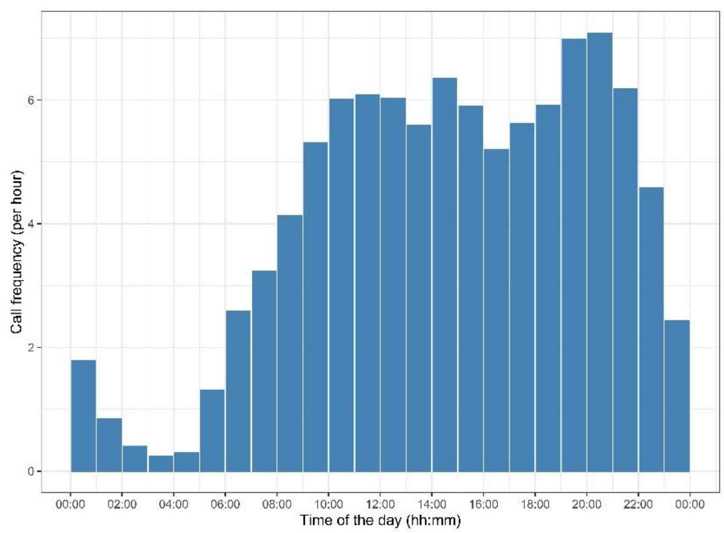


Figure 7-3 Hourly call rate on a typical working day.

To construct the CDR trajectory for a group of agents, we leveraged the set of SGT location and time data generated via the MATSim simulation. Using the distribution of actual call rate (the average number of calls per hour) observed in real CDR data, we produced one-day trajectory data, with the hourly call rate during a typical working day depicted in Figure 7-3. For each 10-second interval within this timeframe, we randomly generated call rates from a normal distribution centred around the population's median call rate (0.053 per hour). By using a Poisson distribution[33], we determined the number of calls for each 10-second interval based on the call rate. Subsequently, we selected a location for each call from the agent's SGT data and generated a call duration ranging from 1 to 60 minutes, following a uniform distribution. Call start and end times were then generated based on these durations. This procedure was repeated for all 10-second intervals within the specified time period. It is noteworthy that to generate the synthetic CDR

[33] Poisson distribution is a discrete probability distribution that can be used to simulate the number of calls when it is known how many calls are made on average per hour during that time (Letkowski, 2012).

data, SGT locations were updated to correspond to the nearest mobile phone tower location (tower locations were extracted from the real CDR data from 2012).

7.2.3. Extracting mobility information

In order to assess the relative accuracy of CDR, GPS and GSM mobility data, we extracted mobility information from the synthetic trajectory. The following paragraphs summarise extracted mobility information and the methods used:

Stay location: To identify potential stay locations, including congestion stay points and potential activity locations from the trajectory data, we utilised a geographic clustering method based on density for GPS and GSM data. Density-Based Spatial Clustering of Applications with Noise (DBSCAN), a classic density-based algorithm, was selected due to its ability to identify clusters of various shapes without the need to specify the number of clusters in advance (Fu et al., 2016). Figure 7-4 shows the conversion process of point density into stop points. We applied a temporal rule within each cluster to remove potential congestion stays before extracting potential activity locations from the cluster data. For the synthetic CDR trajectory generated for a single day, call locations were assumed as potential stay locations. However, to distinguish these potential stay locations from those recorded during travel, we examined the driving distance between the two locations and the time lag between call times. If the time lag between call times exceeded the time required to travel between the two locations by car, the locations were considered as stay locations.

Home and activity location identification*:* With GPS and GSM data, which offer records with a high level of temporal precision, we were able to deduce likely home locations and other activity places from their trajectories. Home locations were identified as places where agents revisited multiple times, with sightings predominantly in the early morning and late evening. In contrast, CDR data recorded very few locations, primarily while calls or texts were made. Distinguishing home and other activity location information from CDR data, especially for agents with a single day of call data following the median call rate distribution, posed challenges. In the absence of location data, we made the assumption that each stay location in the CDR data represented a potential activity location, aligning with the rule proposed by Zilske and Nagel (2014) for generating synthetic CDR data.

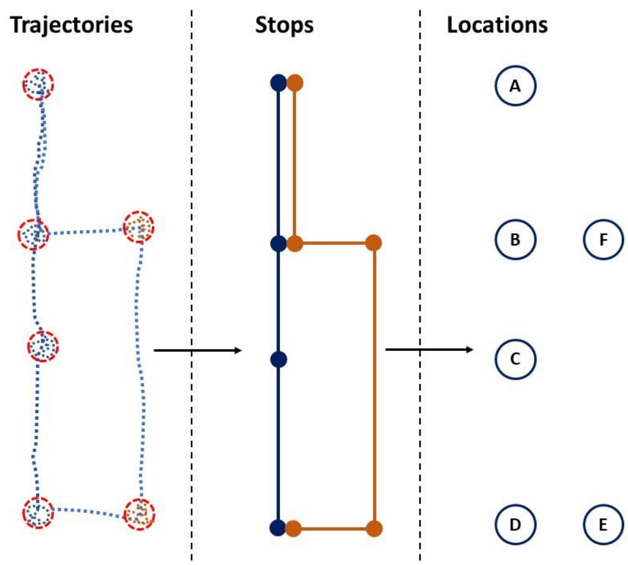


Figure 7-4 Location extraction by density clustering.

Trip attributes: Following the identification of the stay locations, we extracted various trip-related attributes — departure time and trajectory distance, for each trajectory. To extract departure time from each stay location cluster in GPS and GSM trajectories, we sequenced the clustered points based on their observed times. We then examined the first observation and duration within each stay location cluster point, sequentially capturing different start times for repeated activities and their corresponding stay durations. The first observation and corresponding duration in each cluster were used to find the departure time for each stay location to travel for the succeeding stay location cluster. CDR data, however, did not allow for the capture of departure times, as the call and text times recorded in this data reflected the sighting times at activity locations. To calculate the trajectory distance from GPS and GSM data, we sequenced the stay location clusters and calculated the Euclidean distance of trip segments, considering the centroid of each stay location cluster point. The total trajectory distance for each agent was the sum of these individual trip segment distances. For CDR data, we sequenced activity locations based on their sighting times and calculated the total trajectory distance by summing the individual Euclidean trip segment lengths.

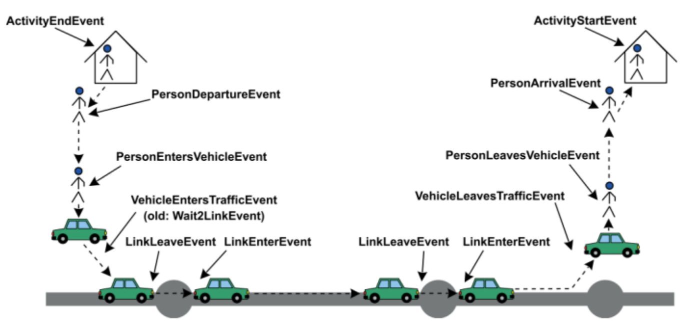
7.2.4. Accuracy assessment

To assess the accuracy of passive mobility data, we compared the statistical and spatial characteristics of individual mobility information extracted from GPS, GSM, and CDR data with the SGT data. We employed correlation statistics to determine the relationship between the synthetic trajectory's distance and the SGT data. Additionally, we assessed the accuracy of the stay locations (potential activity locations). To achieve this, we created grid cells of varying resolutions (e.g., 50m x 50m, 100m x 100m, 200m x 200m, 500m x 500m) within the study region. For each synthetic trajectory, we counted the number of activity locations found within each grid cell. The precision of stay location estimation was evaluated using bivariate analysis.

7.3. Results

7.3.1. Synthetic ground truth (SGT)

The MATSim simulation output mimics real-world agents' activity, departure times, routes, and mode choices. The output is represented in an 'events' file, which comprehensively documents the movements and activities of each agent throughout the simulated day. The various event types encompassed in the MATSim output include "*Activity End Event*", "*Person Departure Event*", *"Person Enters Vehicle Event", "Vehicle Enters Traffic Events", "Link Leaves Event", "Link Enters Event", "Vehicle Leaves Traffic Event", "Person Leaves Vehicle Event", Person Arrival Event",* and *"Activity Start Event"*. A schematic diagram illustrating the full range of events stored in the MATSim simulation output is provided in Figure 7-5 (a). MATSim output offers comprehensive situational information about the agents' actions. For this study, we extracted activity start and end times, and time-stamped link IDs, for 9,704 agents. An example of the extracted event information is demonstrated in Figure 7-5 (b). It should be noted that precise activity location information (exact latitude and longitude) was not saved in the event file; instead, it was derived using the time-stamped link IDs corresponding to each event. The coordinates of activity start, and end points provided insight into the potential home and activity locations of agents, while link coordinates depicted locations during the trips. In total, 20,661 unique locations were identified as potential activity locations, while 2,69,429 locations were identified as en-route point locations, collectively constituting the SGT. Figure 7-6 (a) illustrates SGT data generated from the MATSim simulation for a single agent.



Event type	Time	Person ID	Link ID
actend	23404	134	5108733800005f
left link	23406	134	5108733800005f
left link	23444	134	5108733800005r
left link	23505	134	2965586160006f
left link	23518	134	2638027510020f
left link	23536	134	266421830010f
left link	23555	134	5066951920002f
left link	23570	134	353238520029f
left link	23572	134	353238520030f
left link	23576	134	353238520032f

Figure 7-5 (a) MATSim Events by Axhausen et al. (2016); (b) Extracted SGT information from the event file.

7.3.2. Synthetic trajectory data

Three distinct sets of synthetic trajectory data were generated, each incorporating different levels of noise (detail in section 2.2). For 9,704 agents, a total of 19,333 CDR trajectory points, 4,64,253 GSM points, and 30,728,597 anonymous GPS traces were generated. These trajectory data sets exhibited varying levels of spatial and temporal granularity for an equivalent number of agents. Synthetic GPS, GSM, and CDR data for a representative agent are shown in Figure 7-6 (b to d). Figure 7-6 (b) exhibits the accuracy and precision of GPS data, providing precise location information along the entire route of the trip (assuming the device and location apps were active during the journey). This accuracy was achieved because the synthetic GPS data in this study did not account for urban canyons and topographic effects (Dhaka city's topography is generally flat with fewer concentrations of skyscrapers). Therefore, the temporal and spatial resolution of data from GPS devices was solely influenced by the GPS receiver's type in the vehicle and the satellite network's availability. Figure 7-6 (b) and Figure 7-6 (c) demonstrate that GPS and GSM data offered reasonably accurate information at coarser spatial resolutions, aligning with the SGT. These datasets not only illustrated the travel route but also depicted stopover locations or congested points through point density. Therefore, low-precision GPS data could provide trip-related location updates similar to GSM data. Conversely, CDR data only exclusively presented records at the tower level based on call times, resulting in coarser temporal and spatial precision compared to GPS and GSM data types. The granularity of CDR data depended on tower location density and call rate since it reveals the location of cell towers.

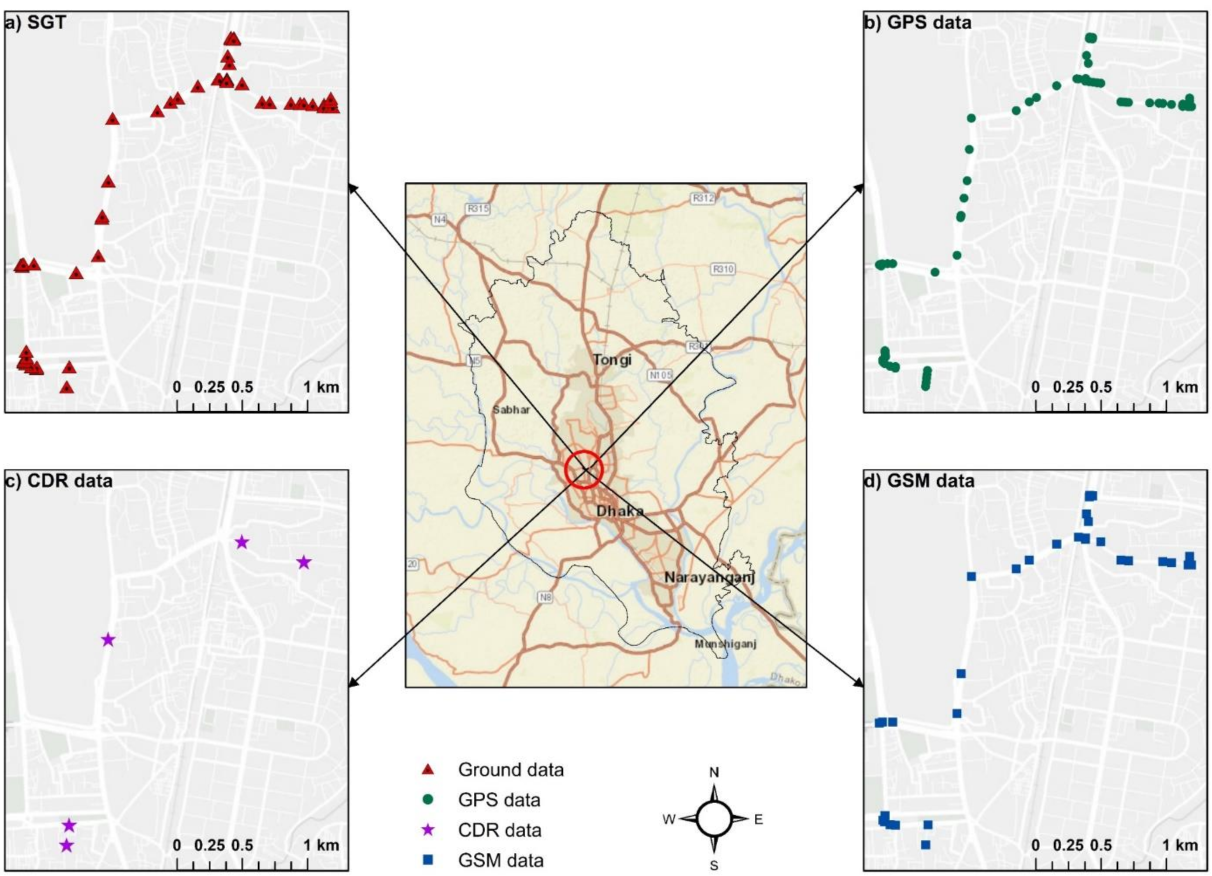


Figure 7-6 Synthetic ground truth and trajectory data (one individual's trajectory).

7.3.3. Analysing the accuracy of different trajectory data

In order to evaluate how well GPS, GSM, and CDR could be useful to extract mobility information, we compared the information extracted from synthetic data with the MATSim simulation-generated SGT data. This comparison encompassed statistical properties of trip-related information and the spatial distribution of location information between synthetic trajectory data and SGT data.

7.3.3.1. Trip related statistics

Departure time information was extracted from the SGT, GPS, and GSM trajectories. The *end time* of each activity served as the departure time for the subsequent activity in the SGT. For instance, the beginning of a trip from home to work involved leaving the house and the end of home-related activity. Therefore, the end of home activity was the time for departure for work activity. The method followed to extract departure time from GPS and GSM trajectory is explained in section 7.2.3. Figure 7-7 illustrates the departure time distribution of GPS, GSM, and SGT trajectories. It is evident that GPS data more accurately captured the variance in departure time during peak hours and better represented agents' departure times for different activities compared to GSM data. However, as shown in Figure 7-7, both GPS and GSM trajectories substantially understated morning departure times compared to the SGT. Conversely, GPS and GSM trajectories overrepresented afternoon departure times. Consequently, the disparity between the SGT and GPS/GSM trajectories was more pronounced during peak hours (e.g., morning peak from 8:00 to 10:00 and afternoon peak from 16:00 to 18:00). This variation may be attributed to the merging of activity or stay clusters near the journey's origin with the nearby congestion clusters, leading to over or underrepresentation during peak hours. Extracting agents' departure times from the one-day synthetic CDR data was challenging. Therefore, the distribution of sighting times (call/text times) extracted from the CDR data is presented in Figure 7-7. The sighting distribution from the one-day CDR data markedly differed from the SGT departure time distribution.

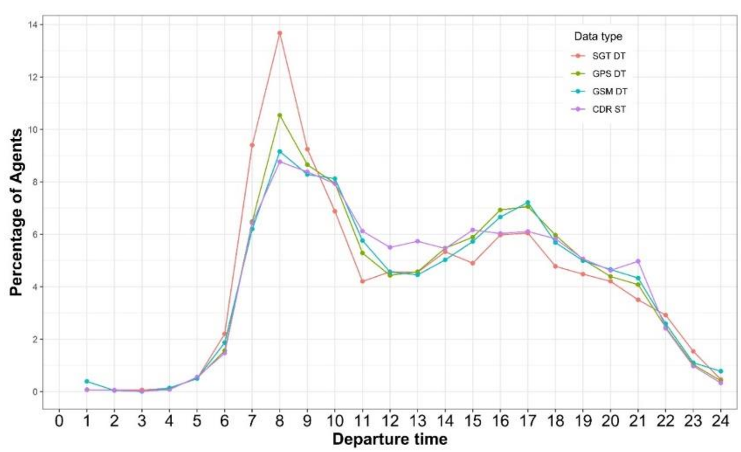


Figure 7-7 Time dimension of GPS, GSM, CDR and SGT trajectory *(DT is the departure time distribution, ST is the sighting time distribution)*.

In addition to departure time, we compared trajectory distances derived from SGT and synthetic passive data. Figure 7-8 depicts the bivariate relationships and correlation coefficients (*r*) between the SGT trajectory and passive trajectories (GPS, GSM, and CDR). GPS trajectories exhibited the highest correlation (r = *0.77*), followed by the GSM trajectories (r = 0.55), and CDR trajectories (r = 0.47). GPS devices provided frequent (10-second intervals), accurate, and precise location and time information, likely contributing to the stronger association between the GPS and the SGT trajectory. In contrast, GSM provided triangulated approximate location data, influenced by the network strength and mobile phone tower density, resulting in a moderate degree of correlation between synthetic GSM trajectories and the SGT trajectory. The lower spatial resolution of the CDR data (attributable to the lack of location records when the mobile phone was not in use) likely accounted for the low correlation coefficient between the CDR trajectory and the SGT. Furthermore, identifying home and possible activity locations using one-day CDR data proved challenging. This difficulty arose because sighting locations in the CDR data could represent locations observed during trips rather than endpoints. Additionally, capturing stay locations from CDR data was influenced by variations in mobile phone usage frequencies among users. Conversely, GPS and GSM data, along with their corresponding trajectories, featured finer temporal and spatial granularity, simplifying the identification of starting, intermediate, and ending locations within the trajectory.

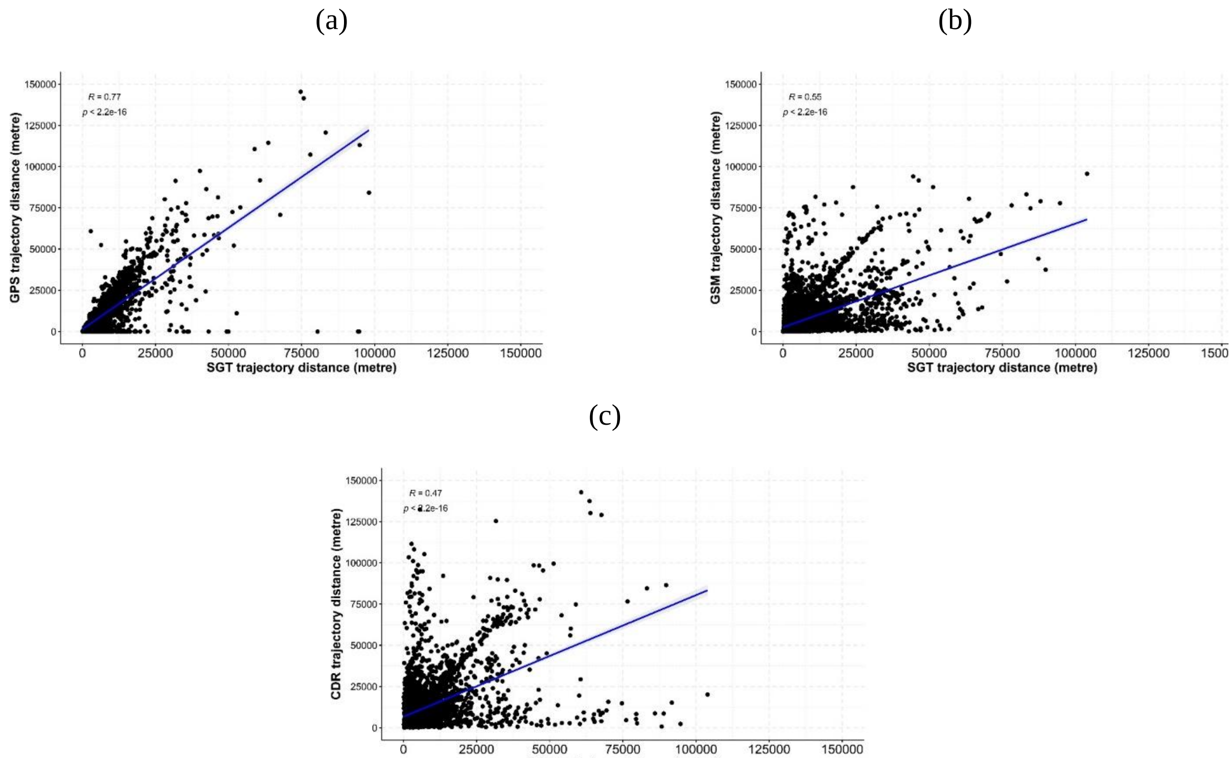


Figure 7-8 Correlation between SGT and synthetic traversed trajectory distance (Euclidean distance).

7.3.3.2. Stay location accuracy

To evaluate the accuracy of the spatial distribution of stay locations, we compared the stay locations extracted from SGT data and synthetic trajectory data. Figure 7-9 illustrates the distribution of stay locations (at a 500m resolution) obtained from the three types of passive data in the Dhaka City Corporation (DCC) region. It is visually evident that the distribution of GPS and GSM data was closely aligned with the distribution of observed activity/stay locations within 500m × 500m grid cells. In contrast, the location data from the CDR dataset conformed to the SGT in central Dhaka, where mobile phone towers were densely concentrated, but exhibited notable discrepancies in other areas, such as the outskirts (e.g., the eastern fringe region). Such discrepancies were more evident when mapping the stay location distributions in the RAJUK area due to variations in tower location density between the DCC area and the surrounding regions. To emphasise these differences through quantitative analysis, the coefficient of determination from the bivariate analysis of SGT and the stay locations from synthetic passive data was compared to assess stay location accuracy, particularly in Dhaka city's RAJUK area.

The results of this analysis at different spatial granularities (i.e., 50m, 100m, 200m, and 500m) are summarised in Table 7-3. At the finest spatial resolution (50m x 50m grid), only 0.07% of the variance in the stay locations from the SGT could be explained by the stay locations extracted from the one-day CDR data. In comparison, GPS data demonstrated the highest explanatory power for stay location distribution at this finer resolution. Indeed, at the finest spatial resolution (50m x 50m grid), the explanatory power of GPS data for stay location distribution was twice that of GSM data (R^2 for SGT vs. GPS was 0.55, whereas R^2 for SGT vs. GSM was 0.27). This trend highlights that, especially at lower resolutions, stay locations extracted from synthetic data exhibited a better alignment with SGT locations. The explicability of GSM stay locations noticeably improved at 100m resolution, roughly doubling from the 50m resolution (R^2 for SGT vs. GSM = 0.478). As the grid size increased, GPS and GSM showed comparable explicability in stay position. However, even at a 500m resolution, CDR data could only explain around 30% of the variation in SGT stay locations, significantly less than the over 80% achieved by GPS and GSM data.

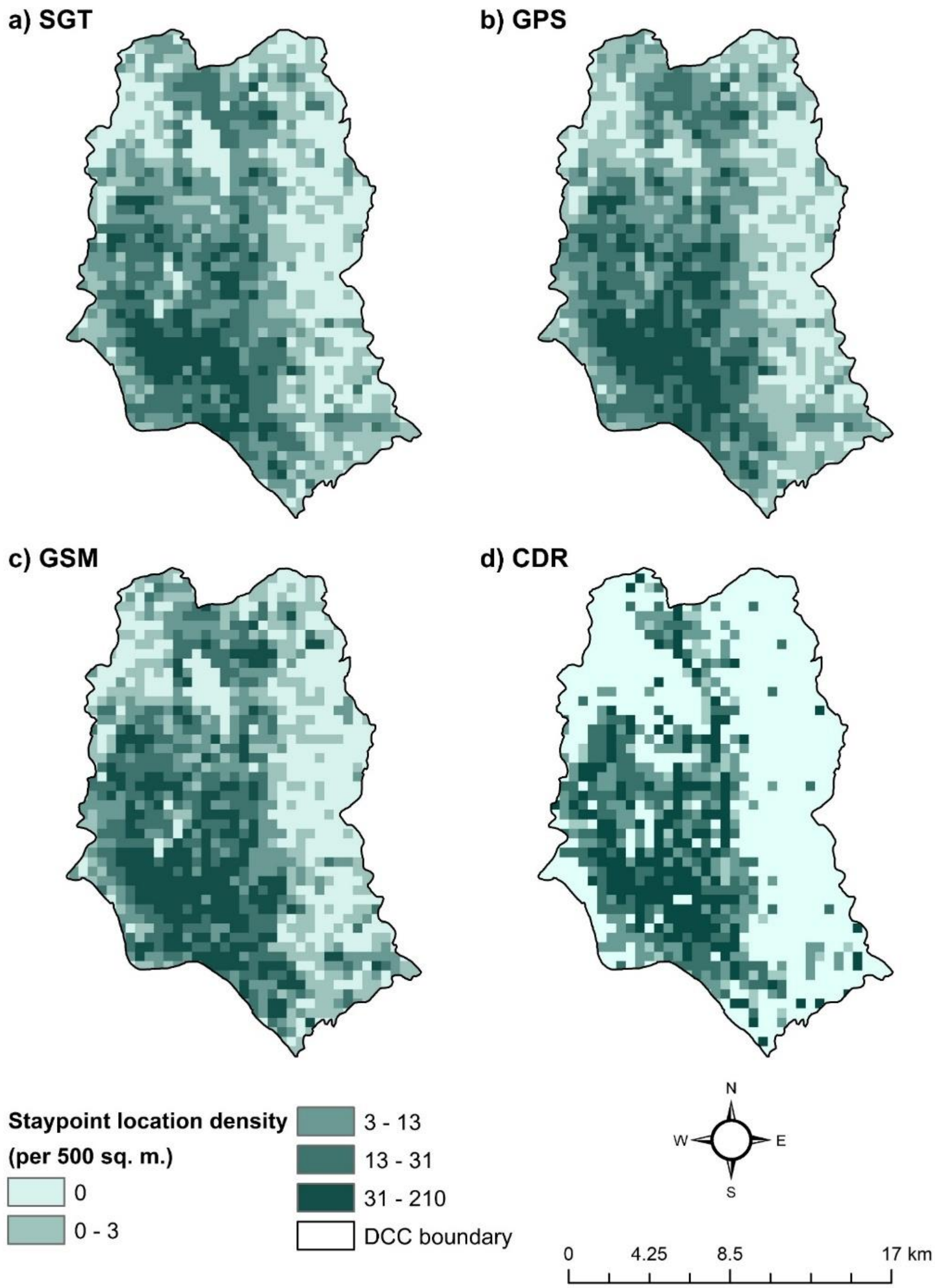


Figure 7-9 Stay-location distribution at the Dhaka City Corporation (DCC) area.

Table 7-3 Summary of bivariate statistical analysis of stay location capturing accuracy between SGT and synthetic trajectory.

Data type	Bivariate component	Cell size resolution			
		50m	100m	200m	500m
CDR	R^2	0.0007	0.0071	0.0583	0.305
GSM	R^2	0.2653	0.478	0.6393	**0.8101**
GPS	R^2	**0.5502**	**0.6476**	**0.7671**	**0.8979**

7.4. Discussion

In this study, we introduced a comprehensive framework for assessing the relative positional and temporal accuracy of mobility information extracted from passive trajectory data, taking into account the limitations of conventional ground-truth data. We utilised MATSim simulation to generate synthetic ground truth (SGT) data. The mobility profile obtained from the MATSim simulation provided the essential inputs, such as link information of individual geocoded activity location, and potential traversed route information, necessary for generating synthetic trajectories using a conventional trajectory generation module. The generation of synthetic data is a crucial step in the field of transport planning, as it allows for the testing of the accuracy and applicability of developed models with reproducible data (Chatterjee and Byun, 2023). Additionally, the generated SGT enabled us to evaluate the accuracy of mobility information from three different passive data sources (GPS, GSM, and CDR) at both the disaggregate and aggregate levels.

The findings of this study shed light on the reliability of passive data in describing mobility profiles during various times of the day, including peak and off-peak hours. Notably, we observed that GPS and GSM data exhibited discrepancies (e.g., either understated or overstated) in the departure time distribution during peak hours (both in the morning and evening), potentially attributed to congestion near activity locations. However, it is important to note that GSM data exhibited a significantly higher measurement error compared to GPS data. The largest difference in departure time distribution between SGT and GPS data was about 3% (with a standard deviation of ± 0.78), while the difference between GSM and SGT departure time distribution was about 4.5% (with a standard deviation of ± 0.98). Interestingly, Bwambale, A. et al. (2019) reported that GSM data exhibited greater accuracy than GPS data when studying departure time choices in southwestern Switzerland. This discrepancy can be mainly attributed to observed GPS data in areas where topographical factors, dense foliage, and human factors (e.g., deactivated location service, and battery power loss) resulted in larger time gaps. Additionally, when compared to GPS and GSM, the time distribution generated from CDR data showed a larger deviation from the SGT time distribution, with a maximum deviation of approximately 5% (with a standard deviation of ± 1.5). This difference can be explained by the fact that one-day CDR data reflects sighting distribution rather than the departure time distribution when a text message or phone call is made (Chen, C. et al., 2016). However, it is worth noting that as mobile phone internet usage (e.g., calls, texts, and browsing) increases, CDR data is evolving towards continuous data, which could eventually achieve a temporal precision comparable to that of GSM data.

The results of the one-to-one comparison between passive data and SGT also revealed that GPS data exhibited the highest level of agreement with the SGT when estimating travelled trajectory distance. Conversely, CDR data demonstrated the lowest level of agreement with the actual trajectory distance. Saarik (2017) further emphasised the error in constructing mobility patterns due to the tower-level resolution of CDR data. Since we calculated the trajectory distance in this study as the sequenced Euclidean distance, accuracy depended on the sequencing and retrieval of the activity or stay locations from passive trajectory data. The sequence of activities in CDR data might not match the actual trajectory because locations are only recorded when the mobile phone is in use. While increased mobile phone usage frequency could potentially improve the sequencing of missing activity locations, the trajectory distance would still be limited to tower positions. Similarly, a weaker association between GSM trajectory distance and SGT, compared to GPS and SGT, can be attributed to the triangulated approximation of stay location information in GSM data.

Furthermore, our assessment of stay location accuracy using GPS, GSM, and CDR data revealed that higher accuracy was associated with lower spatial resolution. Patrick (2016) also emphasised the impact of CDR data resolution on the accuracy of mobility information derived from CDR trajectories. Bwambale, A. et al. (2019) noted that when GSM cell sizes decrease, GSM time lags also reduce. We found at the highest resolution of 50m, both CDR and GSM data exhibited poor performance in explaining the variability of SGT stay locations compared to GPS data. However, at a lower resolution of 500m, both GPS and GSM accounted for over 80% of the variability in stay locations. These findings reinforced the notion that GSM data can effectively serve as a source of mobility information for evaluating activity profiles at coarse spatial resolution. A compromise in the resolution of the scale of analysis has the potential to reduce the measurement error of different passive data.

In addition to device performance, location errors in GPS and GSM data also depend on external factors such as road congestion and the relative distance between activity locations and congested roadways. As a result, deriving activity and travel locations from GPS and GSM data can introduce additional errors beyond positional shift, drift, and discontinuity noise. This may have a significant effect on the methods for location extraction (Fu et al., 2016). Thus, it can be difficult to distinguish short stays (e.g., pick-up or drop-off locations) and activity locations from congestion stays when using GPS and GSM data. This has been particularly problematic in the context of this case study due to the mixed land use and on-street parking facilities on most of the roads in Dhaka. Combining passive trajectory data with secondary data on traffic, weather, and parking could help differentiate various types of stay locations, a potential avenue for future research. Additionally, conducting further research with multi-day panels of GPS and GSM data could aid in better distinguishing activity locations from long stay points resulting from traffic congestion.

7.5. Conclusion

This article introduced a comprehensive four-step framework for assessing the accuracy of mobility information extracted from trajectory data. We employed this framework to generate finely detailed Synthetic Ground Truth (SGT) data and synthetic trajectory data (GPS, GSM, and CDR) infused with realistic noise using MATSim simulation. The use of simulation-based GT allowed us to precisely evaluate multiple passive trajectory data sources and their accuracy in depicting mobility information. Through both visual and statistical analysis, we compared the statistical attributes and accuracy of trip-related factors (e.g., departure time, travel distance, and stay location) extracted from synthetic GPS, GSM, and CDR data with those from SGT. Our findings highlighted that the generated synthetic data had the potential to closely resemble real-world GPS, GSM, and CDR data. Furthermore, when considering the additional positional disturbance, GPS data outperformed GSM and CDR data in terms of deriving departure time, trajectory distance, and activity location information.

Moreover, we empirically demonstrated that the accuracy of passive data is contingent on various assumptions made during their evaluation, such as assuming the mobile device was active throughout the journey. External factors such as congestion and the relative distance from the road to the activity location also play a role. Additionally, the proposed framework offers several key advantages:

- **Importance of Synthetic Ground Truth (SGT)**: This study underscores the significance of having SGT that closely matches the spatial and temporal granularity of passive data sources. We tested its significance with three mainstream passive data. Also, SGT facilitates precise and rigorous comparisons between passive data (from different sources) and a reliable reference in a controlled environment, enabling researchers to assess how accurately these sources capture human mobility patterns and choose the most suitable data source for their specific needs.
- **Addressing challenges with real-life data**: While real-life datasets are essential for accuracy assessment, they pose challenges in distinguishing relative inaccuracies caused by different noise levels and their impact on model uncertainties. The proposed framework would allow researchers to isolate and assess the influence of various types of noise, beyond positional disturbance, validating the stability of model outputs derived from passive data and their sensitivity to error size and extraction assumptions.
- **Alignment with other accuracy assessments**: The accuracy assessment results aligned closely with those from other assessments that use travel diary surveys, census, or other GT sources. This alignment underscores the suitability and significance of the proposed SGT for future accuracy assessment studies. Additionally, the framework's versatility enables the assessment of passive trajectory data accuracy at both the individual and aggregate levels.

Furthermore, this framework provides a foundation for benchmarking models developed with passive data, aiding in the evaluation of various data management solutions. The results from the comparative analyses can help to identify data requirements for different scales of transport planning and modelling. Additionally, accuracy assessments of passive data with appropriate GT have the potential to mitigate challenges related to data collection, processing, and model specification complexities. They can reveal the most suitable passive data source for specific concerns, such as using GSM data instead of GPS data (Ahas et al., 2007a).

It is important to note that the findings deduced from the MATSim-based study using the Dhaka network may not be applicable universally since the accuracy level of the passively generated data is affected by local factors like the topology of the transport network, topographic characteristics of the area (e.g. if the area is flat or mountainous, if there are clusters of high-rise buildings, coverage of the mobile phone network, etc.). For instance, in Dhaka, the mobile phone tower location density notably influenced the accuracy of CDR data in the eastern part of the city. However, this effect and its magnitude may not hold true for cities in developed countries. Considering such contextual factors is imperative when utilising this framework for accuracy assessment. Furthermore, it is essential to have a comprehensive MATSim model for conducting such accuracy assessments, which entails substantial data requirements (e.g., MATSim model inputs such as network, activity plans, behavioural model), and complex technology-dependent simulations (e.g., high-performance computer to run the scenarios). These factors can introduce additional noise during the event file generation stage (e.g., agent teleportation in the absence of a potential route between specified origin and destination in plan files).

Moreover, this study investigated the accuracy of three mainstream passive data sources under a standard positioning (random shift and drift) and temporal noise, future research can explore the sensitivity of these data sources to different noise levels, including variations in shift and drift parameters, tower density, call rate, and location update frequency. Additionally, the impact of noise, arising from external sources such as land use, built environment, and topographic conditions etc., on trajectory data can be explored using the proposed framework. This can be achieved by generating synthetic passive data with added noise from external sources through integrating MATSim event files with external data sources, such as land use maps or weather data. The impact assessment of different types of noise (both internal and external) on trajectory data will also enable to assess the effectiveness of different data processing algorithms while using them to extract trajectory information from passive data sources. Similar investigations can assess the accuracy of other passive data sources, such as smart card data and automatic vehicle location information. Expanding this research to compare model outputs, such as value of travel time (VTT) and other outputs, derived from different passive data sources is also a valuable avenue for exploration. Additionally, conducting a comparative analysis between outputs obtained from MATSim simulation and real GPS, GSM, and CDR data collected from the same sample used for calibrating MATSim can also provide insights into the influence of external factors like density, congestion on assessing the accuracy of mobility information obtained from passive data. While this work focused on checking stay location accuracy at various spatial resolutions (50m to 500m), future research can delve into comparing the relative accuracy of real-time passive trajectory positions and related link/lane level locations by testing various map-matching methods algorithms (Quddus et al., 2005; Bierlaire et al., 2013). Ultimately, the proposed framework holds the promise for generating trajectory data in data-scarce cities and validating them with appropriate GT information to make informed decisions based on validated models developed using passive data. The research community can further enhance and develop new datasets according to their specific requirements.

7.6. Reference

Ahas, R., Aasa, A., Silm, S., Aunap, R., Kalle, H. and Mark, Ü. 2007a. Mobile positioning in space–time behaviour studies: social positioning method experiments in Estonia. *Cartography and Geographic Information Science.* **34**(4), pp.259-273.

Ahas, R., Laineste, J., Aasa, A. and Mark, Ü. 2007b. The spatial accuracy of mobile positioning: some experiences with geographical studies in Estonia. *Location based services and telecartography.* Springer, pp.445-460.

Anda, C., Erath, A. and Fourie, P.J. 2017. Transport modelling in the age of big data. *International Journal of Urban Sciences.* **21**(sup1), pp.19-42.

Axhausen, K.W., Horni, A. and Nagel, K. 2016. *The multi-agent transport simulation MATSim.* Ubiquity Press.

Bar-Gera, H. 2007. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C: Emerging Technologies.* **15**(6), pp.380-391.

Bekhor, S., Dobler, C. and Axhausen, K.W. 2010. Integration of activity-based with agent-based models: an example from the tel aviv model and MATSim. *Arbeitsberichte Verkehrs-und Raumplanung.* **628**.

Bierlaire, M., Chen, J. and Newman, J. 2013. A probabilistic map matching method for smartphone GPS data. *Transportation Research Part C: Emerging Technologies.* **26**, pp.78-98.

Bohte, W. and Maat, K. 2009. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies.* **17**(3), pp.285-297.

Bösche, K., Sellam, T., Pirk, H., Beier, R., Mieth, P. and Manegold, S. 2013. Scalable generation of synthetic GPS traces with real-life data characteristics. In: *Selected Topics in Performance Evaluation and Benchmarking: 4th TPC Technology Conference, TPCTC 2012, Istanbul, Turkey, August 27, 2012, Revised Selected Papers 4*: Springer, pp.140-155.

Bwambale, A., Choudhury, C.F. and Hess, S. 2019. Modelling departure time choice using mobile phone data. *Transportation research part A: policy and practice.* **130**, pp.424-439.

Bwambale, A., Choudhury, C.F., Hess, S. and Iqbal, M.S. 2019. Getting the best of both worlds - a framework for combining disaggregate travel survey data and aggregate mobile phone data for trip generation modelling. In: *15th World Conference on Transport Research India*

Chatterjee, S. and Byun, Y.-C. 2023. A Synthetic Data Generation Technique for Enhancement of Prediction Accuracy of Electric Vehicles Demand. *Sensors.* **23**(2), p594.

Chen, C., Ma, J., Susilo, Y., Liu, Y. and Wang, M. 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies.* **68**, pp.285-299.

Chen, M.Y., Sohn, T., Chmelev, D., Haehnel, D., Hightower, J., Hughes, J., LaMarca, A., Potter, F., Smith, I. and Varshavsky, A. 2006. Practical metropolitan-scale positioning for gsm phones. In: *UbiComp 2006: Ubiquitous Computing: 8th International Conference, UbiComp 2006 Orange County, CA, USA, September 17-21, 2006 Proceedings 8*: Springer, pp.225-242.

Dong, Y. 2022. *Disaggregated Short-term Travel Location Prediction*. Doctor of Philosophy thesis, Imperial College London.

Eagle, N. and Pentland, A. 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing.* **10**, pp.255-268.

Etter, V., Kafsi, M. and Kazemi, E. 2012. Been there, done that: What your mobility traces reveal about your behavior. In: *Mobile data challenge by nokia workshop, in conjunction with int. conf. on pervasive computing*.

Forghani, M., Karimipour, F. and Claramunt, C. 2020. From cellular positioning data to trajectories: Steps towards a more accurate mobility exploration. *Transportation Research Part C: Emerging Technologies.* **117**, p102666.

Fu, Z., Tian, Z., Xu, Y. and Qiao, C. 2016. A two-step clustering approach to extract locations from individual GPS trajectory data. *ISPRS International Journal of Geo-Information.* **5**(10), p166.

Fund, F., Lin, R., Korakis, T. and Panwar, S.S. 2016. How bad is the flat earth assumption? Effect of topography on wireless systems. In: *2016 14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*: IEEE, pp.1-5.

Gong, H., Chen, C., Bialostozky, E. and Lawson, C.T. 2012. A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems.* **36**(2), pp.131-139.

Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H. and Attanucci, J.P. 2013. Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. *Transportation Research Record.* **2343**(1), pp.17-24.

Harding, C., Faghih Imani, A., Srikukenthiran, S., Miller, E.J. and Nurul Habib, K. 2021. Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys. *Transportation.* **48**, pp.2433-2460.
Harrison, G., Grant-Muller, S.M. and Hodgson, F.C. 2020. New and emerging data forms in transportation planning and policy: Opportunities and challenges for “Track and Trace” data. *Transportation Research Part C: Emerging Technologies.* **117**, p102672.
Hasan, A.M., Samsudin, K., Ramli, A.R., Azmir, R. and Ismaeel, S. 2009. A review of navigation systems (integration and algorithms). *Australian journal of basic and applied sciences.* **3**(2), pp.943-959.
Hato, E. 2010. Development of behavioral context addressable loggers in the shell for travel-activity analysis. *Transportation Research Part C: Emerging Technologies.* **18**(1), pp.55-67.
Hemmings, T. and Goves, C. 2017. *Utilising mobile network data for transport modelling: recommendations paper.* United Kingdom.
Hendawi, A., Shen, J., Sabbineni, S.S., Song, Y., Cao, P., Zhang, Z., Krumm, J. and Ali, M. 2020. Noise patterns in GPS trajectories. In: *2020 21st IEEE International Conference on Mobile Data Management (MDM)*: IEEE, pp.178-185.
Huang, H., Cheng, Y. and Weibel, R. 2019. Transport mode detection based on mobile phone network data: A systematic review. *Transportation Research Part C: Emerging Technologies.*
Iqbal, M.S., Choudhury, C.F., Wang, P. and González, M.C. 2014. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies.* **40**, pp.63-74.
Jansen, R., Kovacs, K., Esko, S., Saluveer, E., Sõstra, K., Bengtsson, L., Li, T., Adewole, W.A., Nester, J. and Arai, A. 2021. Guiding principles to maintain public trust in the use of mobile operator data for policy purposes. *Data & Policy.* **3**, pe24.
Janzen, M., Vanhoof, M., Smoreda, Z. and Axhausen, K.W. 2018. Closer to the total? Long-distance travel of French mobile phone users. *Travel Behaviour and Society.* **11**, pp.31-42.
Jones, P. and Stopher, P.R. 2003. *Transport survey quality and innovation.* Emerald Group Publishing Limited.
Kang, X., Liu, L., Zhao, D. and Ma, H. 2020. TraG: A trajectory generation technique for simulating urban crowd mobility. *IEEE Transactions on Industrial Informatics.* **17**(2), pp.820-829.
Kohla, B. and Meschik, M. 2013. Comparing trip diaries with GPS tracking: Results of a comprehensive Austrian study. *Transport survey methods: best practice for decision making.* Emerald Group Publishing Limited, pp.305-320.
Letkowski, J. 2012. Applications of the Poisson probability distribution. In: *Proc. Acad. Business Res. Inst. Conf*, pp.1-11.
Li, Y., Ran, Z., Tsai, L. and Williams, S. 2023. Using call detail records to determine mobility patterns of different socio-demographic groups in the western area of Sierra Leone during early COVID-19 crisis. *Environment and Planning B: Urban Analytics and City Science.* p23998083231158377.
Liao, L., Fox, D. and Kautz, H. 2007. Extracting places and activities from gps traces using hierarchical conditional random fields. *The International Journal of Robotics Research.* **26**(1), pp.119-134.
Liu, J., Li, J., Li, W. and Wu, J. 2016. Rethinking big data: A review on the data quality and usage issues. *ISPRS journal of photogrammetry and remote sensing.* **115**, pp.134-142.
Merry, K. and Bettinger, P. 2019. Smartphone GPS accuracy study in an urban environment. *PloS one.* **14**(7), pe0219890.
Oliveira, M.G.S., Vovsha, P., Wolf, J., Birotker, Y., Givon, D. and Paasche, J. 2011. Global positioning System–Assisted Prompted recall household travel survey to support development of advanced travel model in Jerusalem, Israel. *Transportation research record.* **2246**(1), pp.16-23.
Pan, Y., Sun, Q., Yang, M., Darzi, A., Zhao, G., Kabiri, A., Xiong, C. and Zhang, L. 2023. Residency and worker status identification based on mobile device location data. *Transportation Research Part C: Emerging Technologies.* **146**, p103956.

Pappalardo, L., Ferres, L., Sacasa, M., Cattuto, C. and Bravo, L. 2020. An individual-level ground truth dataset for home location detection. *arXiv preprint arXiv:2010.08814.*

Patrick, A.J. 2016. *Using Simulation to Investigate the Accuracy of Mobility Data Derived from Call Detail Records*. MSc thesis, University of Leeds.

Paul, B.S. and Rimer, S. 2013. Wireless sensor node placement due to power loss effects from surrounding vegetation. In: *Quality, Reliability, Security and Robustness in Heterogeneous Networks: 9th International Conference, QShine 2013, Greader Noida, India, January 11-12, 2013, Revised Selected Papers 9*: Springer, pp.915-927.

Pelletier, M.-P., Trépanier, M. and Morency, C. 2011. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies.* **19**(4), pp.557-568.

Quddus, M.A., Noland, R.B. and Ochieng, W.Y. 2005. Validation of map matching algorithms using high precision positioning with GPS. *The Journal of Navigation.* **58**(2), pp.257-271.

Ratti, C., Frenchman, D., Pulselli, R.M. and Williams, S. 2006. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and planning B: Planning and design.* **33**(5), pp.727-748.

Renfro, B.A., Stein, M., Reed, E.B. and Villalba, E.J. 2022. *An analysis of Global Positioning System Standard Positioning Service performance for 2020*.

Saarik, A. 2017. *Trajectory Reconstruction and Mobility Pattern Analysis Based on Call Detail Record Data*. MSc thesis, University of Tartu

Schipperijn, J., Kerr, J., Duncan, S., Madsen, T., Klinker, C.D. and Troelsen, J. 2014. Dynamic accuracy of GPS receivers for use in health research: a novel method to assess GPS accuracy in real-world settings. *Frontiers in public health.* **2**, p21.

Shen, L. and Stopher, P.R. 2014. Review of GPS travel survey and GPS data-processing methods. *Transport reviews.* **34**(3), pp.316-334.

Smith, D.M., Clarke, G.P. and Harland, K. 2009. Improving the synthetic data generation process in spatial microsimulation models. *Environment and Planning A.* **41**(5), pp.1251-1268.

Song, X., Long, Y., Zhang, L., Rossiter, D.G., Liu, F. and Jiang, W. 2020. Spatial accuracy evaluation for mobile phone location data with consideration of geographical context. *IEEE Access.* **8**, pp.221176-221190.

Steenbruggen, J., Borzacchiello, M.T., Nijkamp, P. and Scholten, H. 2013. Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. *GeoJournal.* **78**, pp.223-243.

Sternfeld, B., Jiang, S.F., Picchi, T., Chasan-Taber, L., Ainsworth, B. and Quesenberry, C.P. 2012. Evaluation of a cell phone-based physical activity diary. *Medicine and science in sports and exercise.* **44**(3), pp.487-495.

Stopher, P., FitzGerald, C. and Xu, M. 2007. Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation.* **34**, pp.723-741.

Toole, J.L., Colak, S., Sturt, B., Alexander, L.P., Evsukoff, A. and González, M.C. 2015. The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies.* **58**, pp.162-177.

Tsoleridis, P., Choudhury, C.F. and Hess, S. 2022. Deriving transport appraisal values from emerging revealed preference data. *Transportation Research Part A: Policy and Practice.* **165**, pp.225-245.

Vanhoof, M., Lee, C. and Smoreda, Z. 2020. Performance and sensitivities of home detection on mobile phone data. *Big Data Meets Survey Science: A Collection of Innovative Methods.* pp.245-271.

Varshavsky, A., De Lara, E., Hightower, J., LaMarca, A. and Otsason, V. 2007. GSM indoor localization. *Pervasive and mobile computing.* **3**(6), pp.698-720.

von Mörner, M. 2017. Application of call detail records-chances and obstacles. *Transportation research procedia.* **25**, pp.2233-2241.

Wang, Z., He, S.Y. and Leung, Y. 2018. Applying mobile phone data to travel behaviour research: A literature review. *Travel Behaviour and Society.* **11**, pp.141-155.

Wolf, J., Bricka, S., Ashby, T. and Gorugantua, C. 2004. Advances in the application of GPS to household travel surveys. In: *National Household Travel Survey Conference, Washington DC*.
Wu, R., Luo, G., Shao, J., Tian, L. and Peng, C. 2018. Location prediction on trajectory data: A review. *Big data mining and analytics.* **1**(2), pp.108-127.
Yang, Y., Xiong, C., Zhuo, J. and Cai, M. 2021. Detecting home and work locations from mobile phone cellular signaling data. *Mobile Information Systems.* **2021**, pp.1-13.
Zannat, K.E. and Choudhury, C.F. 2019. Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science.* **99**(4), pp.601-619.
Zhang, G., Rui, X., Poslad, S., Song, X., Fan, Y. and Wu, B. 2020. A method for the estimation of finely-grained temporal spatial human population density distributions based on cell phone call detail records. *Remote Sensing.* **12**(16), p2572.
Zilske, M. and Nagel, K. 2014. Studying the accuracy of demand generation from mobile phone trajectories with synthetic data. *Procedia Computer Science.* **32**, pp.802-807.
Zmud, J., Lee-Gosselin, M., Munizaga, M. and Carrasco, J.A. 2013. *Transport survey methods: Best practice for decision making.* Emerald Group Publishing Limited.

Appendix E

Supplementary Table 7-1 Strengths and weaknesses of different passive data sources.

Data Type	**Accuracy**	**Precision**	**Spatial data density**	**Temporal data density**	**Population Coverage**	**Semantic Info**	**Level of Bias**
GPS (Smart Phone)	High	High	High	High	High	No	High
Mobile Phone (CDR)	Medium	Low	Low	Low	High	No	Medium
WiFi	High Medium	Medium	Medium	Medium	Medium	No	Medium
Social Media	High	High	Low	Low	Low Medium	Yes	High
Travel Card	High	High	Low	Low	Medium	No	Low

Source: (Dong, 2022)

Chapter 8 Discussion and conclusions

8.1. Summary

This thesis has looked at the theme of transport planning in the Global South with a focus on data and behavioural modelling issues has been presented in this thesis. The introduction (Chapter 1) detailed a number of areas that this thesis concentrated on. These areas were chosen among the current behavioural research issues from the Global South. Some of them were relevant to situations from the Global South where there were comparable data and modelling issues and others might apply to a universal context. Some of these issues be of concern in the Global North, however, their applicability was evaluated in this thesis with a focus on the Global South as their importance predominated in the south in comprehending the behavioural complexity. Examples include the sensitivity to activity start time, variability of satiation for mandatory and discretionary activities, etc. This part of the thesis revisits these themes, establishes connections between the various chapters of the thesis, and provides a summary of the research's conclusions.

8.1.1. Data issues prevailing in the Global South

Data of reasonable sample size and acceptable accuracy levels is an essential pre-requisite for developing robust behavioural models. Developing reliable behavioural models is therefore challenging (if not impossible) in the context of the Global South due to a lack of trustworthy data sources containing necessary information required by the modellers. Data related issues are prevailing in both conventional (e.g., manual survey-based RP data) and emerging third-party data source (e.g., emerging data from Google Maps API). Due to the lack of information in the conventional data sources required by the modeller, model developed with unreliable data may result in erroneous prediction of travel demand. To address the challenges associated with conventional data sources, empirical research has advocated exploring emerging data sources. Certain studies from Global South, albeit limited in number, have endeavoured to construct advanced mode choice models, aiming to mitigate data scarcity concerns by leveraging alternative data types such as SP data. However, few studies have attempted to develop other types of choice behaviour such as departure time choice model as it requires large-scale data at a finer spatial and temporal granularity. Some of the required information for developing departure time choice model is more often unavailable in conventional survey-based RP data. For example, travel time for unchosen alternatives is typically unavailable in the RP data. SP data have the problem of a smaller sample size. Moreover, Preferred departure/arrival time, etc. is usually unavailable in the RP data. While in the Global North, emerging data sources (Google Maps API, Microsoft Bing map) are used to get such information, in the context of the Global South, they are very often insufficient as a source of information on their own (detail in Chapter 2). The widely used "best guess" traffic model from the Google Maps API, assumed to offer more reliable travel time estimates for Global North, more often fails to provide accurate information for various origin-destination pairs and alternative timeframes in the context of Global South. In this thesis, using secondary data sources, advanced discrete choice models of car commuters' time-of-day choice were modelled by selecting Dhaka, Bangladesh as a case. Travel diary data from commuters using personal vehicles or ride-hailing services (957 and 934, respectively) for trips from their homes to their workplaces was used. As the Google Maps API cannot be a stand-alone source of travel time information in the context of Dhaka, a new method to estimate the travel time for the complete range of alternative time periods using stated travel times and the three traffic models of Google Maps API was suggested in this work (detail in Chapter 2). The proposed framework of this study quantified how the level-of-service attributes (e.g., travel time), socio-demographic characteristics (e.g., type of job, income, etc.), and situational constraints (e.g., schedule delay, activity duration) affected the departure time preferences of car commuters while accounting for the data limitations. Additionally, the proposed method included two distinct statistical distributions for office workers and self-employed individuals acknowledging the high level of heterogeneity between and within each group, as opposed to assuming a constant value for a particular market segment or a generic statistical distribution (Chapter 2). The specification used in this research was able to demonstrate that instead of relying on ad-hoc assumptions, advanced discrete choice models can produce behaviourally realistic results, and emerging data sources have the potential to address data related challenges related to using traditional data sources for advanced modelling.

Moreover, the existing travel survey data lacks information on users' responses to potential future modes of transport such as BRT, MRT, and subway. Hence, the majority of the travel demand prediction for the inaugural mode relies on ad-hoc assumptions. For example, it may be assumed that the preference for BRT and MRT, as well as the sensitivity to travel time and cost after the implementation of these services, will remain consistent with the patterns observed in existing PT usage. However, such an assumption may not hold true in the context of the Global South due to the radical differences in the level of service between the BRT/MRT and the existing PT. Chapter 6 of this thesis presents the development of a joint RP-SP mode choice model for simulating BRT demand in the context of Dhaka, Bangladesh. The developed model considered users' responses to currently available mode (RP data) and hypothetical scenarios (SP data). On one hand, the adopted approach addressed the data challenges arising from the lack of information on inaugural modes in the conventional travel surveys by using SP data. Besides, the unavailability of unchosen alternative modes in RP data was resolved by utilising the Google Maps API. In the absence of travel time data for the other motorised modes (PT, autorickshaw, or motorcycle) in Google Maps, a congestion matrix was developed based on car travel times for various TAZ pairs. Mode-specific travel times at free-flow speed were then used to estimate the travel times of other alternative motorised modes. Further, the joint RP-SP model overcame the hypothetical bias of the SP-only model. The estimated model yielded more reliable results and demonstrated the relative preference for the inaugural mode. In this model, common parameters (travel time and travel cost) as well as specific parameters unique to the RP and SP were computed using the joint model (ASCs and distance). The estimation result of the value of travel time (VTT) was also consistent with similar studies by Rahman et al. (2020). This adopted approach of inferring unavailable information from emerging data sources, rather than solely relying on RP data, would enable the development of other behavioural models such as destination and route choice models.

In addition to the issues associated with conventional data, this research also focused on challenges in behavioural modelling while employing passively generated big data. One of the main issues raised in the literature was the lack of appropriate ground truth data needed to verify the accuracy of passive data and benchmark the model estimated using widely available data sources. For instance, location data from passive GPS and GSM data were at the individual geo-location level, whereas widely used ground truth from trip diary survey data provided TAZ-level location information. Due to the dearth of suitable ground truth data, studies have resorted to using census data, traffic counts, or other alternative passive sources as substitutes for ground truth when conducting parallel surveys to obtain true ground truth for passive data is unfeasible due to privacy concerns. However, employing ground truth data that does not align with the temporal and spatial resolution of the relevant passive data may result in inaccurate model outputs. Very few studies have tried to measure the accuracy of the passive data and attempted to benchmark the model developed using these ubiquitous data sources. In Chapter 7, a simulation-based approach is summarised to generate synthetic ground truth (SGT) data that had similar temporal and spatial granularity with the regularly used passively generated big data sources, namely GPS, CDR, and GSM. Through a comparative analysis (SGT and passive trajectory data), the accuracy of inferring the underlying human mobility patterns from each of these three selected data sources was evaluated. Mobility information was extracted from the trajectory data by identifying and distinguishing potential stay locations and measuring trip-related attributes such as departure time, trajectory distance, etc. Additionally, Chapter 7 has brought attention to the challenges (e.g., stationary conditions due to congestion wrongly labelled as stay locations) of extracting mobility information from trajectory data. It may be noted that though this is a particular problem in megacities of the Global South, the problem can also be prevalent in the case of cities in other parts of the world. The statistical and spatial characteristics of individual mobility information extracted from the GPS, GSM, and CDR data were compared with the SGT data to assess the accuracy of passive trajectory data, both at the disaggregate and aggregate levels. The results shed light on the accuracy level of the outputs derived from these data sources and the extent of bias to the SGT. When real ground truth data is unavailable, synthetic ground truth can serve as a means to gauge the precision of passive trajectory data and assess the influence of specific noise on mobility data derived from that particular source. Understanding the accuracy of these data sources in comparison with ground truth could enable researchers, policymakers, and practitioners to select the best data source for their requirements, ultimately leading to more effective and efficient decision-making processes in a variety of disciplines.

Eventually, optimising investment in the transport sector in the Global South requires enhancing the understanding of travel behaviour and improving the accuracy of travel demand predictions. This entails minimising errors stemming from (unreliable) data used for modelling due to the absence of information in the readily available data sources. While emerging data sources hold promise for complementing conventional data, it remains imperative to validate their accuracy and make required adjustments prior to their incorporation into behavioural modelling and policymaking frameworks. This approach will contribute to reducing errors in demand prediction and facilitate the development of more robust planning tools tailored to the needs of the Global South.

8.1.2. Modelling behavioural complexity

To capture behavioural complexity, the field of behavioural modelling has been constantly evolving. Some studies attempted to concentrate on a single aspect of travel and activity behaviour, whereas others tried to use a joint model to incorporate multiple aspects into a single framework. In this research, state-of-the-art models were assessed to see if they can incorporate additional behavioural complexity such as correlation, heterogeneity, and endogeneity which has been either omitted or not explored in a sufficient level of detail in contemporary studies. Though the complexity modelled could be applicable universally, in this thesis, we analysed them empirically from the perspective of the Global South.

The departure time choice of travellers was traditionally modelled using parameters associated with three dimensions of trip choice – outbound, return, and duration. However, departure time decisions for outbound and return legs, and the corresponding activity durations, are not independent decisions but rather interrelated in most cases. Due to extended work hours, a rigid work schedule, an office-based work culture, a lack of facilities for remote work, and other factors prevailing in the Global South, consideration of this correlation and its implication in behavioural modelling is crucial while modelling time-of-day-choice focusing on the Global South. However, none of the earlier departure time choice models specifically looked into how this possible correlation might affect model results. In this thesis, a utility structure was suggested with a novel polynomial functional form of alternative specific constants (ASCs), which captures this correlation in a joint (outbound and return) departure time choice model, to fill in this gap in the existing literature. In Chapter 3 of this thesis, joint departure time choice models developed using the proposed polynomial functional form and the trigonometric specification proposed by Ben-Akiva and Abou-Zeid (2013) are incorporated. Compared to the trigonometric model, the proposed model showed a better model fit with fewer parameter estimates. The findings highlighted that if correlations persist across various activity and travel dimensions, disregarding those correlations may lead to inaccurate predictions of demand. Therefore, for peak spreading policy development, taking into account the correlation between outbound departure time and duration in the time-of-day choice model could aid in improving departure time prediction in situations where such correlation was present.

Similar to departure time, appropriate treatment of the time dimension such as activity start time is an important prerequisite to determining activity demand or forecasting the travel demand accurately. Two specific time dimensions of activity participation are highlighted in the literature: activity time use and activity timing. Among these two dimensions, the latter one has received less attention in activity-based travel behavioural research. In chapter 4 of this thesis, the empirical investigation of the effect of activity start time in the activity type and duration model is incorporated. A 24-h travel diary survey data (70,756 respondents) from the Dhaka Metropolitan Region (RAJUK area), Bangladesh was used for the purpose of this investigation. The activity type and duration models were estimated using different multiple discrete-continuous (MDC) models — 1) MDCEV model (without activity start time) with non-negativity and a binding budget constraint, 2) MDCEV model accounting for utility differences based on activity start time with non-negativity and a binding budget constraint, 3) bounded MDCEV model accounting for utility differences depending on activity start time accommodating time of day specific time budget constraints along with non-negativity and a binding budget constraint. In reality, the decision maker encounters time of day specific budget constraints while accommodating multiple activities within a limited time period. For instance, a strict inequality in the distribution of time throughout the day may be a result of a rigid workplace schedule, a family duty, a deadline, opening and closing time of activity locations, or an individual's preference or ability. The decision-maker must allocate their time to alternatives while maintaining such inequal time bounds (for certain activities) keeping the daily time budget's general equality constraint in mind. Chapter 4 summarises the detail of the adopted modelling framework to jointly model activity type, start time and duration using a fairly simple scenario of morning and afternoon activity bifurcation. The results from the classical MDCEV and bounded MDCEV (both considering activity start time) model revealed that depending on the activity's start time (morning/afternoon), utility and satiation associated with choice alternatives also varied across different market shares (male, unemployed, have access to a motorised vehicle, high-income households, etc.). Despite the fact that the budget should not be the same at different time of the day, the bounded MDCEV model (with the inequality constraint based on the starting time of the activity) had a marginally better prediction (though not significant) than the model without a time-based constraint. One of the likely reasons for the lack of improvement in prediction accuracy was the coarse bifurcation of the day (24-hour and 12-hour), as none of the respondents reached the upper bound by spending the entire time on one activity. Even if the prediction accuracy of joint modelling with inequal budget constraint does not show its importance, it can still reduce forecasting errors, ensuring that none of the respondents are overpredicted and preventing exceeding the maximum allowable budget. From a policy perspective, the proposed model using activity start time information jointly with activity type and duration provided rich situational information about activity and travel patterns which can be used for activity timing in an agent-based simulation model.

While it is agreed that travel demand is driven by activity needs, the relationship between activities and travel is far from uniform. In order to better understand the factors influencing complex activity-travel choices and establish causal links between them, an appropriate modelling framework is required because the decision to engage in an activity is intricately linked to and correlated with several circumstances. In the literature, to improve the prediction accuracy of activity demand, activity related choice dimensions have been modelled using different modelling framework. However, to model time use decisions, the MDCEV model has received wider attention as it offers corner solution and accommodate satiation. Although time use decisions can also be modelled using the FMNL model (considering time use as fraction), MDCEV has predominated the current activity research due to its flexibility to accommodate satiation (diminishing marginal utility) associated with activity time use preference. But regardless of the activity category, requirements, or performance circumstances, all activities may not lead to satiation. Satiation can be person-specific and context-specific. Examples include watching a football match (for a football fan), working overtime (for a workaholic or for an individual facing financial challenges), or working extra hours before a deadline. Chapter 5 of this thesis presents a comparative analysis of activity choice and time use modelling using both the FMNL and MDCEV models. It explores whether different specification of satiation and correlation across alternatives in time use modelling improves the prediction accuracy of activity duration or enables better capturing the homogeneity within the group while accounting for heterogeneity across the group. A detailed multi-day travel diary survey data accumulated for seven consecutive days (working days and weekends) from 170 residents of the Greater Concepción, Chile has been used in this regard. Activity durations predicted by FMNL and MDCEV were examined for both in-home and out-of-home activities. Both selected modelling framework has the potential to incorporate multiple continuous components of the time-use decision. Results revealed that compared to the MDCEV model, FMNL had a higher prediction accuracy and the potential to parsimoniously recover the average time use. Additionally, FMNL model can be used effectively with improved prediction accuracy for the activities that were observed to be engaged in by the majority of the sample and required little time investment. However, the logit formulation of FMNL implicitly considers the diminishing marginal utility attributed to the non-additive nature of the framework. Unlike the MDCEV model, the FMNL model does not possess the capability to estimate whether a particular socio-demographic group experiences faster or slower satiation in their time usage. An improvement in the MDCEV model fit was evident after incorporating heterogeneity in baseline preferences and satiation (latent class MDCEV framework) and capturing correlations among the alternatives (MDCNEV model). However, accommodating correlation in MDCNEV model improved prediction accuracy of average discrete probability which was reflected by the improvement in prediction accuracy of activity duration if participated. However, MDCNEV improved the discrete probability in the expense of predicting average time use. Similarly, for the selected sample, rather than heterogeneity in satiation, heterogeneity in baseline preferences significantly contributed to the enhancement of model fit within the latent class MDCEV framework. In Chapter 5, the baseline utility and satiation effect varied among three distinct classes (21.6%, 48.9%, and 21.5%), with one class more likely to participate in in-home activities, another group in out-of-home mandatory work activities, and the third group more inclined toward out-of-home discretionary activities. While distinctive variations in activity participation were discernible across the classes, the differences in satiation were relatively subtle. However, the latent class FMNL failed to capture the heterogeneity across different classes and did not demonstrate any improvement in the model fit. Therefore, it is worthwhile to investigate for a given sample profile whether the consideration of translated utility explicitly accommodating satiation or the implicit accommodation of satiation within the logit framework enhances the prediction accuracy of activity duration. Furthermore, the correlation among alternatives and the impact of random heterogeneity in time use modelling vary depending on the chosen modelling framework, underscoring the importance of selecting an appropriate framework to accurately predict activity time use. The results could guide transport modellers in specifying satiation, correlation among the alternatives while modelling time use preference to predict activity duration and formulating optimum market strategies that can address their travel needs.

Due to socio-economic and cultural differences between the Global South and north, models developed focusing on the Global North cannot be directly applied in the Global South. In this thesis, three behavioural dynamics have been investigated and incorporated in the state-of-the art models: strong correlation between activity duration and time-of-day choice, time-dependent utility of activity, and the role of satiation, correlation and random heterogeneity in activity time use modelling. The lessons from the investigation suggest adapting behavioural modelling frameworks with contextual dynamics prevailing in the Global South. This adaptation may not escalate the data requirement since all investigations utilised conventional travel and activity diary survey data. This adaptation will enable to enhance model's resilience in accommodating both current and emerging issues of behavioural complexities from Global South.

8.1.3. Simulation model in the Global South and potentiality

The advent of advanced information and telecommunications technologies and their application to transportation systems have expanded the range of options for computationally expensive modelling and simulation in understanding behavioural dynamics. Agent-based microsimulation models, widely used stochastic simulation models, are frequently used to simulate urban traffic patterns, activity-travel patterns, impact assessments, service performance evaluations, accessibility evaluations, location decisions, and many other things. The majority of agent-based microsimulation models, however, are concentrated on cases from the Global North. Using those cases as a reference case for the Global South would be problematic as those modelling framework ignore the behavioural and contextual dynamics prevailing in the Global South. Very few studies addressed the difficulties in developing agent-based microsimulation models, which concentrated on the Global South, into practice. To address issues regarding travel demand prediction in the Global South, in Chapter 6 a demand forecasting model for an inaugural BRT is summarised. In an agent-based microsimulation tool, a mode choice model representing the preference for emerging transport modes was implemented. The model was developed in the setting of Dhaka, the capital of Bangladesh and one of South Asia's and the world's fastest-growing megacities. This was done using the multi-agent, activity-based, transport demand simulator MATSim. Using behaviour models in MATSim to represent mode choice in the presence of the proposed BRT was the focus of the MATSim implementation in the context of Dhaka. The modelling framework also incorporated a variety of data sources (including stated-preference data) for calibrating the mode choice and other MATSim components to accurately replicate urban travel patterns. After calibration, MATSim was used to simulate various BRT access scenarios and evaluate the sensitivity of the outputs to various modelling assumptions (sensitivity to travel time, cost, and availability of access mode). The simulation findings demonstrated how substantially BRT travel demand was impacted by the marginal utility of travel time, travel cost, and pricing structure. The planners would be able to maximise the ridership of the proposed BRT with the help of the direct model outputs given in Chapter 6 of this thesis, but the calibrated simulator would also be helpful for the assessment of other emerging transport modes in the context of Dhaka in the future. Emerging simulation models offer a comprehensive approach to addressing complex transport planning issues in the Global South, ranging from congestion reduction to demand prediction. Rather than relying solely on traditional models and adopting a conservative approach to planning decisions, simulation frameworks can be utilised to better understand behavioural complexities in the Global South.

8.2. Objective and contribution

Objective 1: Investigating how emerging third-party data (e.g. Google Maps API) and passively generated mobility data (e.g. GPS, GSM and CDR) can enhance the limited traditional data available in the Global South, to facilitate the application of advanced modelling frameworks. This objective was met with the development of Chapter 2, Chapter 6 (partly) and Chapter 7. In Chapter 2, the empirical investigation of the Dhaka setting brought to light the difficulties in developing robust behavioural models when both traditional and emerging data sources have limitations. In Chapter 6, the development of joint RP-SP mode choice model showed how both conventional and emerging data sources can be used to understand complex behavioural issues under uncertainty. In Chapter 7, it is demonstrated how (synthetic) ground truth produced by agent-based simulation can be used to evaluate the accuracy of mobility information derived from new data sources.

The potential to address some of the data issues prevalent in the Global South (e.g., accounting for uncertainties associated with Google Maps API data, and unobserved preferred departure times) is highlighted in Chapter 2 and Chapter 6. <u>The dissertation thus contributes to understanding issues of data scarcity and their significance in developing robust behavioural models in the context of the Global</u>

South. It also suggests that, with certain corrections or adjustments, emerging data sources possess the potential to tackle data-related challenges in utilising traditional data sources for advanced modelling.

Objective 2: Evaluating the existing state-of-the-art model in addressing emerging issues of transport planning dynamics in the context of the Global South This objective was met with a comparative analysis between the proposed model and the state-of-the-art models. State-of-the-art models developed focusing on developed countries may be significant in the context of the Global North, but behavioural models from the Global South must take into account the prevailing behavioural dynamics in the context. Chapters 3, 4, and 5 of this thesis cover three interrelated behavioural dynamics — correlation, time-dependent constraint, and role of satiation. To address the correlation between outbound and return, in a joint departure time choice model, the proposed polynomial functional form was compared with the trigonometric formulation proposed by Ben-Akiva and Abou-Zeid (2013). To highlight the importance of activity start time in the activity type and time use model, the bounded MDCEV model having activity start time in the model was compared with the classical MDCEV model by Bhat (2008) with and without activity time-wise segregation. Also, in investigating the role of satiation, the FMNL model was compared with the MDCEV model. The findings of the comparative analysis emphasised the imperative for a more nuanced understanding of the behavioural dynamics prevailing in the Global South. The results highlighted the need to go beyond the simple implementation of state-of-the-art model. Rather, it became clear that these models need to be improved and expanded to better account for the unique behavioural dynamics under consideration. This contribution sheds light on the importance of tailoring modelling approaches to align with the unique context of the Global South.

Objective 3: Developing agent-based micro-simulation models for forecasting demand for Bus Rapid Transit (BRT). This objective was met in Chapter 6 which focused on an application combining a discrete choice model with a simulation model to capture complex behavioural issues. In the context of Dhaka, demand for the inaugural BRT under different circumstances was simulated and the sensitivity of the outputs to different modelling assumptions (travel time, cost, and access mode) was tested. Therefore, this thesis contributes to extending the methodological framework of travel behaviour research by advocating for the integration of complementary models (discrete choice model and ABM) for a more comprehensive assessment of future transport scenarios involving new modes.

8.3. Future research directions

In addition to the contributions already mentioned, each chapter and the associated study suggest several directions for further investigation. Following is a summary of those potential research questions and future research directions.

The modelling framework proposed in Chapter 2 shows a procedure to use secondary data (Google Maps API and RP data) to model the departure time choice of car commuters in Dhaka, Bangladesh. Moreover, Chapter 6 showed methods to infer travel time for other motorised modes and measured the travel cost only based on network distance and standardised fare (per km cost). In addition to Dhaka, it is crucial to conduct similar evaluations of travel time using Google Maps API in other cities or locations from the Global South. This will strengthen the application of emerging data sources to complement the necessary information required by modellers for the development of advanced choice models. However, in this thesis, the accuracy of the measured travel cost was not evaluated with a similar temporal granularity as travel time. This was due to the lack of information about the vehicle type and the type of driver, as well as users stating fares for public transport, paratransit, and other motorised modes. In future research, departure time choice models that take into account users of PT and other modes as well as the sensitivity to travel costs can be estimated using primary data or appropriate supplementary data. Additionally, the robustness of the developed departure time choice models was assessed from the demand perspective using arbitrary forecasting analysis. In future research, it will be important to integrate the departure time choice model into a traffic assignment model to analyse its efficiency considering the supply side as well. Implementing these demand models in a traffic assignment model will enable the quantification of the effectiveness of proposed policies such as congestion pricing and reduced public transport fares. Similarly, the agent-based simulation framework proposed in Chapter 7 investigates the accuracy of three different mainstream passive data sources considering a standard form of positioning disturbance. In addition to addressing positioning disturbances, future research could explore the sensitivity of these data sources to various levels of noise. This could involve investigating how different factors such as shift and drift parameters, tower density, call rate, and location update frequency affect the accuracy of capturing human mobility information. Following a methodology similar to that employed in this thesis, the accuracy of the synthetic trajectory data, incorporating noise from external sources such as density, land use, weather, and congestion, can be verified against the ground truth generated by MATSim. To do so, synthetic trajectory data could be generated incorporating various types of noise by combining weather, land use, built environment, and topographic data with the MATSim event file. Also, to delineate the noise generated from the MATSim simulation, a comparative analysis between output obtained from simulation and real GPS, GSM, and CDR data collected from the same sample can be performed. The focus of this work was restricted to evaluating the accuracy of the stay location at various spatial resolutions (ranging from 50m to 500m). However, by experimenting with various map-matching approaches, future research can focus on evaluating the relative accuracy of real-time passive trajectory positions and relevant link/lane level locations. The research in this thesis can be further expanded to compare model outputs estimated from various passive data sources, such as VTT and other outputs. Additionally, along with the chosen sources, data from smartphone apps, smart card data, and other passive data sources offering travel-related information can be included in a similar comparative analysis in future studies to check their accuracies in inferring spatio-temporal trajectories. Eventually, cities with varying topography, land use, density, and built environmental characteristics should be selected from both the Global South and north regions. This will allow for testing the robustness of the proposed framework in assessing the accuracy of passive data.

Besides the data-related challenges, Chapter 3 to Chapter 5 comparatively have assessed the state-of-the-art models in capturing the behavioural dynamics such as correlation, effects of satiation, and activity start time predominated in the Global South. However, the proposed joint departure time choice model (capturing the correlation between departure time and duration requirement) was focused on car commuters and their joint trip for work activities. However, a detailed investigation is necessary to get additional insights about non-commute trips and non-home-based trips. It may be noted that the departure time choice for non-commute trips is more complex as there may be more flexibility associated with the choice of activity destination and mode. In such cases potentially warrants a joint model for departure time, destination, and mode. Also, the proposed framework ignored the potential correlation between adjacent departure times to retain simplicity for practical implication. Future research can focus on estimating correlations of alternatives by using a more robust modelling framework, such as a cross-nested logit model or mixed MNL model. Besides, activity timing consideration in activity behavioural modelling has been elaborated in Chapter 4, which attempted to shed light on the importance of the consideration of activity start time in the activity type and duration choice model. A very straightforward scenario of morning and afternoon activity bifurcation (with appropriate bound on each time choice) was used to evaluate how activity start time affected activity decision and satiation. Activity start time can be broken down into finer scales in future studies (e.g., morning peak, morning off-peak, evening peak, evening off-peak, etc.), along with specific and appropriate upper (and lower) bounds for activity types based on institutional opening and closing time. Furthermore, the authors did not classify in-home activities in detail to examine the significance of activity timing preference. Additionally, it is crucial to consider the time interval between two consecutive activities in activity time preference and time usage decision in future studies, especially for multiple participation requirements of an activity. The proposed model was unable to capture the correlation between activities carried out in the morning and afternoon by applying a potential error structure because this study was conducted using 24-hour survey data. Panel data can be used to further expand the current model by capturing the correlation among the alternatives as well as the inter-intra heterogeneity connected to utility and satiation of activity timing and time use decision. Furthermore, the output from the bounded MDCEV model, which provides time-dependent utility for different socio-demographic groups, can be utilised for activity timing in an agent-based simulation model. This approach would be valuable in comprehending the influence of individuals' activity decisions on a macro scale, such as at the neighbourhood or city level.

The effects of satiation in activity participation decisions, investigated by comparing results from the FMNL and MDCEV models focusing on data from Concepción, Chile, are summarised in Chapter 5. Though the study exhibited that for certain activities (observed to be participated by the major share of the sample or required limited time investment) FMNL was appropriate for modelling time use decisions compared to the MDCEV model. However, this failed to propose a framework capable of addressing all the complexities discussed in the chapter. To achieve this, it is crucial to conduct similar comparative analyses with other candidate models to determine which one provides the most accurate predictions of duration with minimal errors. Moreover, this thesis was limited to the comparative analysis of FMNL and MDCEV models using conventional data sources. However, the satiation effect can be further investigated using passive data rather than relying on traditional travel diary survey data to determine how sensitive the satiation effects were brought on by short and long-stay activities (which are frequently missed in travel diary survey data) and their corresponding obligations to be completed in a week. For instance, it is possible to predict likely time use and associated satiety of respondents for both short and longer-duration activities utilising smart card information or high-resolution GPS data. Moreover, Chapter 6 presents an agent-based simulation approach to capture the influence of certain time and cost sensitivity and access to the relevant facilities in enumerating the potential demand of an inaugural BRT route network. The models used in the ABM simulation were however quite simplistic combining mode choice scenarios. Departure time choice of different market share and activity start time specific utility specifications are yet to be implemented. Additionally, agent-to-agent communication (including that between family members and with friends and coworkers) was not taken into account in the current version of the model. In order to examine the effects of such dynamic interaction occurring inside their social network of agents, the current model can be updated to incorporate dynamic interaction. Moreover, this study was limited to only one large investment that is currently being implemented in Dhaka. Similar studies are required for the other investments that are in the pipeline, such as the MRT, flyover, and expressway, as well as combinations of these initiatives in order to assess the combined effectiveness of such investments prior to the implementation of the project. Future research using longitudinal data can assess how the current or preferred choice will be impacted by the change in land use, service quality, habits, awareness, evolving attitude, culture, social norms, and values after the successful BRT and MRT operations in Dhaka.

Overall, the research provided in this thesis shows how emerging data sources might address some of the data-related challenges (both in conventional and emerging data) inherent in developing advanced behavioural models and transport planning tools, particularly those prevalent in the Global South. To complement traditional data sources, more studies using emerging data sources are required. Additionally, through the comparative analysis, this thesis emphasises the need for further extending the cutting-edge models to take into account the behavioural dynamics that are prevailing in the Global South. All the comparative analysis results from this thesis highlight that contextual adaptation is required before employing models that use the Global North as a reference point.

8.4. References

Ben-Akiva, M. and Abou-Zeid, M. 2013. Methodological issues in modelling time-of-travel preferences. *Transportmetrica A: Transport Science.* **9**(9), pp.846-859.

Bhat, C.R. 2008. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological.* **42**(3), pp.274-303.

Dong, Y. 2022. *Disaggregated Short-term Travel Location Prediction*. Doctor of Philosophy thesis, Imperial College London.

Rahman, F.I., Bari, M., Islam, M. and Joyanto, T.P. 2020. ANALYSIS OF MODE CHOICE BEHAVIOR AND VALUE OF TIME IN DHAKA CITY, BANGLADESH. *International Journal for Traffic & Transport Engineering.* **10**(2).